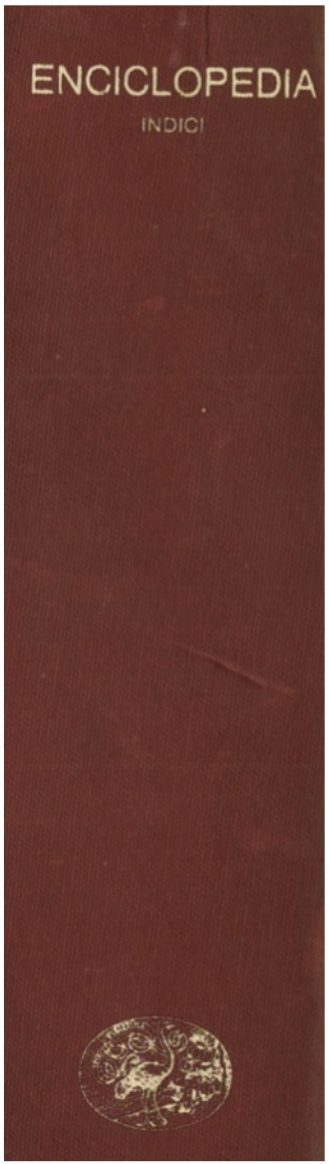
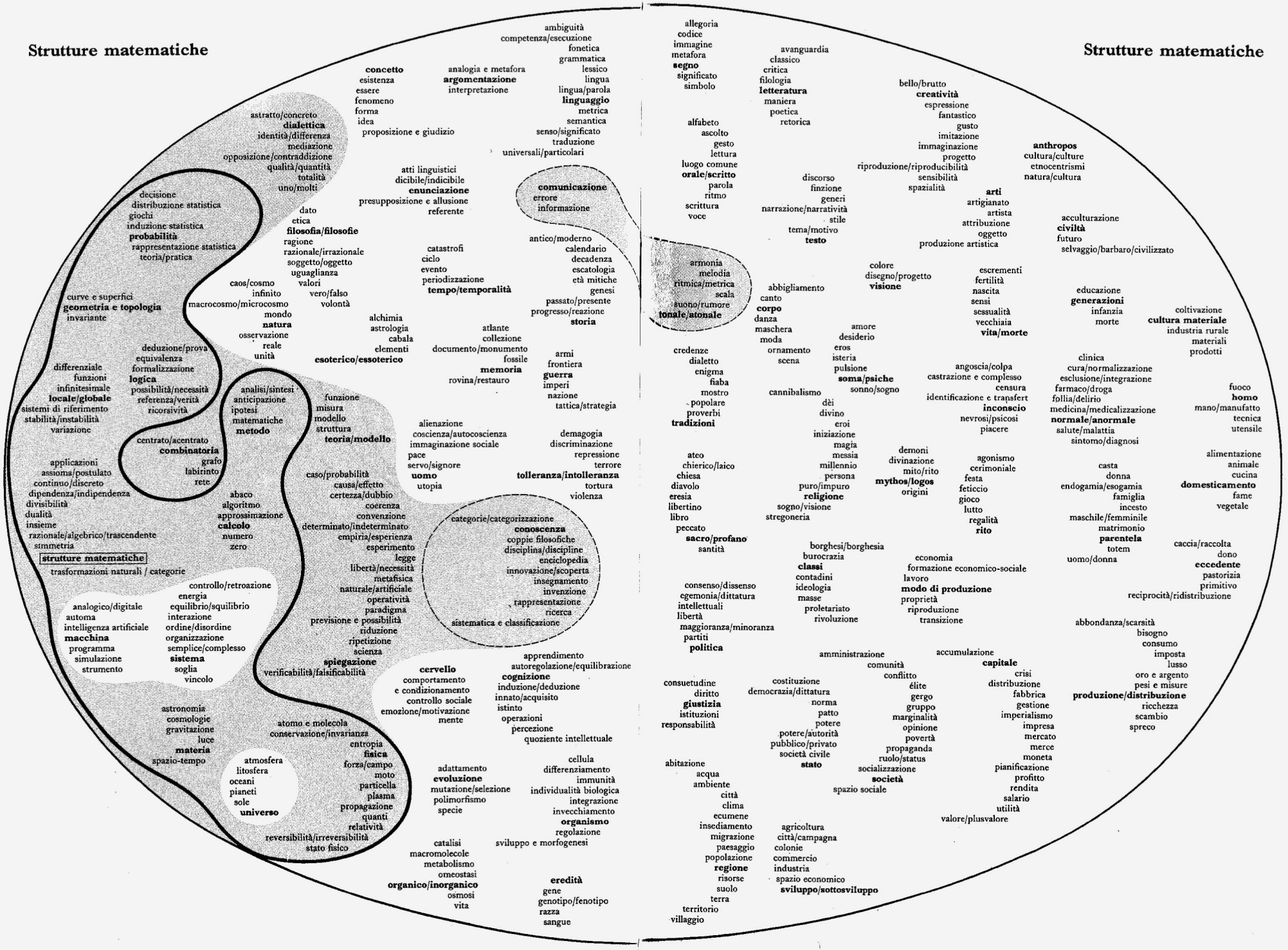


ENCICLOPEDIA EINAUDI [1982]

STRUTTURE MATEMATICHE

Massimo Galuzzi - **STRUTTURE MATEMATICHE** pag.5

Jacques Le Goff - **APPLICAZIONI** pag.11
ASSIOMA/POSTULATO pag.33
CONTINUO/DISCRETO pag.43
DIPENDENZA/INDIPENDENZA pag.70
DIVISIBILITÁ pag.93
Israil Moiseevič Gel'fand e Jurij Ivanovič Manin - **DUALITÁ** pag.105
Jurij Ivanovič Manin - **INSIEME** pag.132
RAZIONALE/ALGEBRICO/TRASCENDENTE pag.149
Israil Moiseevič Gel'fand e Jurij Ivanovič Manin - **SIMMETRIA** pag.163
Jurij Ivanovič Manin - **STRUTTURE MATEMATICHE** pag.177
Massimo Galuzzi - **TRASFORMAZIONI NATURALI/CATEGORIE** pag.195



Strutture matematiche

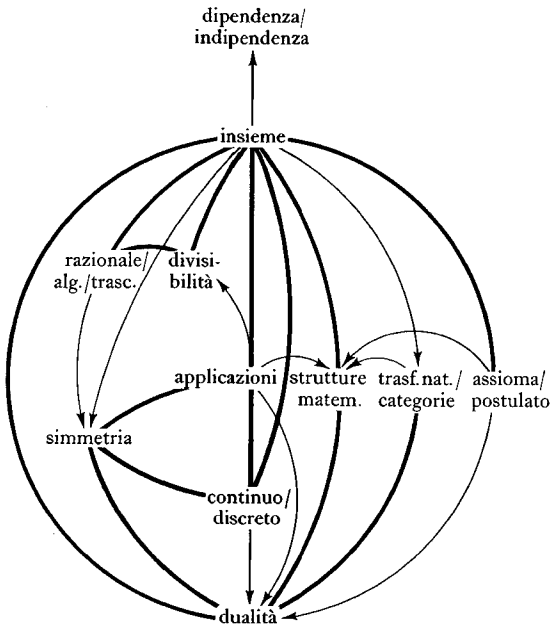
	anthropos	argomentazione	arti	calcolo	capitale	cervello	civiltà	classi	cognizione	combinatoria	comunicazione	concetto	conoscenza	corpo	creatività	cultura mat.
applicazioni	·	·	·	6	4	·	·	1	2	4	5	3	·	2	3	4
assioma/postulato	3	·	·	6	4	·	·	·	3	6	4	6	6	·	5	·
continuo/discreto	4	·	3	6	·	3	·	3	·	4	7	4	6	·	4	·
dipendenza/indipendenza	·	·	·	·	6	2	4	4	3	4	5	4	4	·	4	·
divisibilità	·	·	·	5	·	·	·	·	·	·	·	·	·	·	·	·
dualità	·	·	·	4	·	3	·	·	4	4	·	2	6	·	·	·
insieme	3	·	·	7	3	5	·	3	5	8	6	7	7	3	4	3
razionale/algeb./trascend.	·	·	·	7	·	·	·	·	·	2	·	4	·	·	·	·
simmetria	2	·	3	·	3	2	·	·	·	4	4	3	6	2	3	·
strutture matematiche	·	·	·	5	·	·	·	·	6	5	·	2	3	·	3	·
trasform. nat. / categorie	4	·	2	·	5	·	1	3	4	4	2	4	5	4	6	·

	memoria	metodo	modo di produz.	mythos/logos	natura	normale/anormale	orale/scritto	organico/inorgan.	organismo	parentela	politica	probabilità	produzione/distrib.	regione	religione	rito
applicazioni	·	5	·	·	2	·	·	·	3	·	·	6	·	2	·	·
assioma/postulato	·	6	2	2	5	1	2	6	2	·	2	5	2	2	·	·
continuo/discreto	·	5	·	2	5	·	5	6	4	5	·	7	·	·	3	2
dipendenza/indipendenza	·	3	4	·	4	5	2	4	4	3	·	5	2	4	4	3
divisibilità	·	·	·	·	5	·	·	·	·	·	·	·	·	·	·	·
dualità	·	3	·	·	·	·	3	3	2	2	·	3	·	·	·	·
insieme	4	6	2	2	5	3	3	6	4	5	·	6	·	6	2	1
razionale/algeb./trascend.	·	5	·	·	·	·	·	·	3	·	·	3	·	·	·	·
simmetria	1	3	·	3	·	·	·	5	5	5	·	5	3	3	·	·
strutture matematiche	·	6	·	·	3	·	·	·	3	·	·	·	·	·	·	·
trasform. nat. / categorie	3	4	4	3	·	2	·	5	4	4	·	3	2	3	2	2

	dialettica	domesticam.	eccedente	enunciazione	eredità	esoterico/essoter.	evoluzione	filosofia/filosofie	fisica	generazioni	geometria e topol.	giustizia	guerra	homo	inconscio	letteratura	linguaggio	locale/globale	logica	macchina	materia
applicazioni	6	·	·	2	·	·	·	3	7	·	8	·	·	·	3	2	3	8	5	6	5
assioma/postulato	6	·	·	2	·	2	·	6	8	·	6	5	4	·	2	3	5	6	8	3	6
continuo/discreto	7	·	·	·	2	·	5	3	7	·	6	·	2	·	1	3	4	8	5	6	6
dipendenza/indipendenza	5	·	·	·	2	·	4	3	6	·	4	·	3	·	3	·	4	5	2	4	5
divisibilità	4	·	·	·	·	·	·	·	2	·	4	·	·	·	·	·	·	5	·	·	3
dualità	4	·	·	·	·	4	·	2	·	·	6	·	·	·	·	·	2	5	2	·	·
insieme	7	2	·	·	3	·	·	4	8	·	8	4	2	3	3	2	7	9	9	4	3
razionale/algeb./trascend.	5	·	·	·	·	·	·	·	2	·	8	·	·	·	·	·	·	8	3	·	·
simmetria	5	·	3	·	·	·	·	2	7	·	3	2	·	·	4	3	·	5	2	3	7
strutture matematiche	6	·	·	·	·	·	·	4	7	·	7	·	·	·	·	·	2	8	4	4	6
trasform. nat. / categorie	7	5	1	4	·	2	5	2	5	·	6	4	1	2	1	3	4	4	·	1	3

	sacro/profano	segno	sistema	società	soma/psiche	spiegazione	stato	storia	strutture matem.	sviluppo/sottosvil.	tempo/temporalità	teoria/modello	testo	tolleranza/intoll.	tonale/atonale	tradizioni	universo	uomo	visione	vita/morte
applicazioni	1	1	4	2	·	7	·	·	9	·	3	4	·	·	1	·	3	2	·	·
assioma/postulato	·	5	4	3	·	7	3	5	8	3	5	6	3	3	3	2	4	3	·	·
continuo/discreto	·	4	5	2	·	4	·	5	8	·	5	3	·	·	3	1	2	·	·	3
dipendenza/indipendenza	1	4	5	3	1	6	3	·	5	6	3	4	·	·	·	·	·	2	·	·
divisibilità	·	·	·	·	·	·	2	·	5	·	·	·	·	·	·	·	·	·	·	·
dualità	·	·	·	·	2	·	2	2	6	2	·	2	·	·	·	·	·	3	·	·
insieme	·	4	6	4	·	6	2	·	9	·	5	4	3	·	2	·	2	·	·	·
razionale/algeb./trascend.	·	·	2	·	·	·	·	·	7	·	·	2	·	·	·	·	·	·	·	·
simmetria	·	3	4	·	·	4	·	4	6	3	5	4	·	·	2	2	4	3	3	4
strutture matematiche	·	·	3	·	·	·	·	·	8	·	·	·	·	·	·	·	4	·	·	·
trasform. nat. / categorie	2	4	7	3	2	6	·	1	7	·	4	4	1	·	·	2	3	4	2	·

	strutture matematiche	insieme	dualità	simmetria	applicazioni	continuo/discreto	razionale/algebrico/trascendente	divisibilità	assioma/postulato	trasformazioni naturali / categorie	dipendenza/indipendenza
strutture matematiche	·	4	7	4	·	·	4	3	·	·	2
insieme	7	·	7	5	6	5	6	5	5	6	5
dualità	4	3	·	5	·	·	·	·	·	4	·
simmetria	2	·	5	·	3	4	·	·	·	·	3
applicazioni	6	5	7	5	·	5	1	5	·	4	3
continuo/discreto	4	6	5	4	2	·	4	4	·	1	·
razionale/algebrico/trascendente	4	2	4	5	·	·	·	6	·	·	·
divisibilità	·	2	·	1	·	·	4	·	·	·	·
assioma/postulato	6	6	5	·	·	1	4	2	·	4	4
trasformazioni naturali / categorie	6	·	5	·	·	2	·	·	·	·	2
dipendenza/indipendenza	2	·	·	·	2	·	4	·	·	·	·



wood Cliffs

The

Strutture matematiche

Applicazioni, Assioma/postulato, Continuo/discreto, Dipendenza/indipendenza, Divisibilità, Dualità, Insieme, Razionale/algebrico/trascendente, Simmetria, Strutture matematiche, Trasformazioni naturali / categorie

L'«identità» degli oggetti matematici nell'evoluzione temporale, l'inseribilità in discorsi sempre piú ampi, che è stata oggetto del commento a ♦Geometria e topologia♦, certo ha colpito l'attenzione di molti pensatori. Un esempio fra i molti è dato da Pierre Duhem: «Le proposizioni che costituiscono le scienze unicamente matematiche sono, al massimo grado, verità universalmente accettate; la precisione del linguaggio, il rigore dei procedimenti dimostrativi, non lasciano spazio a divergenze durature tra le concezioni dei vari geometri; attraverso i secoli, le dottrine si sviluppano in continuo progresso, senza che le nuove conquiste facciano perdere nulla dei domini acquisiti in precedenza».

Ancora piú chiaro è Duhem ove mette a contrasto fisica e geometria: «La fisica non progredisce come la geometria che aggiunge nuove proposizioni definitive e indiscutibili a quelle che già possedeva, progredisce invece perché l'esperienza produce di continuo nuovi disaccordi tra leggi e fatti e perché, incessantemente, i fisici ritoccano e modificano le leggi per poter rappresentare i fatti in modo piú esatto».

Questa identità degli enti matematici ha certo riscontro in fatti reali. Ma essa è un estremo dell'opposizione identità/differenza e si vedranno ora alcuni esempi, avendo cura di rilevare i contrasti che coesistono con le concordanze.

Si è spesso osservato come il metodo assiomatico sia per eccellenza il metodo della matematica (cfr. l'articolo ♦Assioma/postulato♦, per esempio. Si pongano anche a confronto i due articoli «Funzioni» e ♦Applicazioni♦ e si vedrà nel secondo caso come venga organizzata una descrizione assiomatica di molte delle questioni trattate nel primo articolo). Questo metodo ha indubitabilmente origini antiche: certo ben prima di Euclide si sono avute costruzioni teoriche organizzate in modo deduttivo, a partire da poche premesse assunte inizialmente. Tuttavia è con gli *Elementi* che si dispone di un testo articolato e sistematico con il quale si può istituire un confronto fecondo.

Ecco la prima proposizione euclidea:

PROPOSIZIONE I. *Su una retta terminata data costruire un triangolo equilatero.*

La costruzione (e con ciò la dimostrazione della sua esistenza) consiste nel tracciare il cerchio BCD, con centro A e raggio AB, il cerchio ACE, con centro B e raggio BA e nel congiungere il punto C (una delle due intersezioni dei cerchi) con A e B. Poiché $AC = AB$ e $BC = AB$ «cose che sono uguali a una stessa sono uguali anche tra loro» e perciò $AC = BC = AB$ (fig. 1).

Si è osservato fin dall'antichità come l'esistenza del punto C che dà una delle intersezioni dei cerchi non segua dagli assiomi e dai postulati precedenti. Il testo

euclideo va quindi integrato. Un nuovo moderno «Euclide emendato» sarebbe per molti aspetti assai piú ricco e sofisticato dell'originale se si dovesse tener conto del gran numero di osservazioni, come la precedente, che si sono accumulate nei secoli. Ma non è tanto questa differenza che preme qui segnalare. Se si supplisce l'opportuno postulato dell'intersezione dei cerchi, si osserva poi immediatamente che la dimostrazione euclidea diviene completamente rigorosa. Essa ha inoltre caratteristiche d'immediatezza e di semplicità tali da renderla praticamente insostituibile.

Non solo quindi viene riconosciuto negli *Elementi* lo stesso nostro ideale dimostrativo perfino nei luoghi ove occorrono integrazioni, ma anche qui, e in molti altri casi, ci si trova d'accordo nel proporre come le piú naturali le stesse dimostrazioni, pur con le opportune modifiche.

Ma questa identità sembra quasi costituire uno sfondo per mettere in maggior rilievo molte e fondamentali differenze che riguardano principalmente la natura degli enti geometrici. Certo gli enti geometrici degli *Elementi* non sono enti fisici, ma «enti ideali» in senso pieno; ciò non toglie che essi organizzino una descrizione che nel suo complesso va collegata allo spazio della nostra esperienza.

I concetti moderni di «modello» e di «struttura», cosí importanti e fecondi, sono privi di senso se visti in connessione con gli *Elementi*. Questi ultimi, nella misura in cui sono pensati descrivere l'unica geometria possibile, sono per ciò stesso l'unico modello e l'unica struttura geometrica, il che svuota di significato questi concetti, se applicati in questo caso.

Inoltre, sebbene si sia debitori alla scienza greca di molti dei concetti che ora si ritrovano nell'area descritta da ♦simmetria♦, vi sono molte sottili differenze, soprattutto a livello dimostrativo. Non vi è intanto una diretta derivazione del concetto espresso in greco da σύμμετρος 'commensurato, proporzionale' (cfr. l'articolo ♦Razionale/algebrico/trascendente♦), aggettivo usato da Euclide [X, 1] per indicare grandezze che ammettono una misura comune («Σύμμετρα μεγέθη λέγεται τὰ τῷ αὐτῷ μέτρῳ μετρύμενα», ecc.), cioè grandezze commensurabili nel nostro linguaggio. Ma al di là dello slittamento di significato dei termini, è nei metodi dimostrativi stessi che si notano profonde differenze.

Con molta evidenza la cosa si nota in Apollonio. Ogni qual volta nel libro I del suo trattato appaiono le «sezioni opposte», cioè i due rami di una stessa iperbole (come nella prop. 59 del I libro, per fare un esempio tra i piú tipici), le considerazioni di ♦simmetria♦ sono ridotte in modo da rendere le dimostrazioni assai strane per un lettore che non abbia familiarità con la matematica greca. Nella proposizione citata, dopo aver costruito una sezione (ossia un ramo dell'iperbole) Apollonio procede a ricostruire i dati e poi la sezione opposta, anziché ricostruire immediatamente la sezione cercata usando la simmetria. E non è che un esempio fra innumerevoli altri.

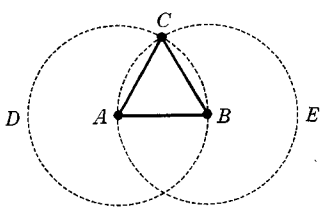


Figura 1.
Costruzione di un triangolo equilatero.

Tornando ad Euclide, nel libro V degli *Elementi* ove è esposta la teoria delle proporzioni, il maggiore strumento dimostrativo della matematica classica, si ha la seguente proposizione (11):

PROPOSIZIONE. *I rapporti che sono uguali a un medesimo rapporto sono uguali fra loro, cioè da $A:B=C:D$ e $C:D=E:F$ si deduce $A:B=E:F$.*

Che significa questo? Conferma ancora una volta che la proporzione $A:B=C:D$ va pensata come un tutto unico, e non come affermante l'uguaglianza di rapporti (in caso contrario al teorema si potrebbe sostituire la prima delle *nozioni comuni*: «Cose uguali a una stessa cosa sono uguali tra loro»).

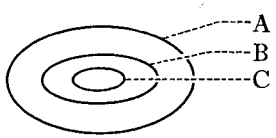
Quante implicazioni questo possa avere sulla considerazione di figure simmetriche rispetto a una data retta, per esempio, è evidente. La considerazione di rapporti «a sinistra» e «a destra» non conduce automaticamente a proporzioni proprio perché, almeno di primo acchito, non è lecito questo modo di intendere le proporzioni. Chi ha familiarità con la matematica greca sa bene quanta pazienza sia talora necessaria per liberarsi di certi «pregiudizi» moderni e intendere anche il senso degli sviluppi dimostrativi in luoghi ove la simmetria sembrerebbe vanificare ogni necessità dimostrativa.

Ora certo i *Fondamenti della geometria* (*Grundlagen der Geometrie*, 1899), la «carta dell'assiomatica moderna» a giudizio di Bourbaki, sono pur sempre una presentazione della geometria erede per molti aspetti della tradizione euclidea, ma la direzione del successivo programma di Hilbert, volto a formare delle concezioni generali dei formalismi matematici, quali giochi con i testi simbolici non richiedenti, in linea di principio, alcuna interpretazione, per usare le parole di Manin, è irriducibile ai termini della matematica greca.

Il confronto con la scienza greca è profondamente significativo anche per un altro aspetto disciplinare.

Secondo un giudizio universalmente acquisito la fondazione della logica si deve ad Aristotele.

Ora, quando si legge, ad esempio, nei *Primi analitici* [25b, 37-39] che «se A si predica di ogni B, e se B si predica di ogni C, è necessario che A venga predicato di ogni C», ci si trova certo su un terreno familiare a tutti coloro che hanno una qualche esperienza matematica e l'idea di una figura come



(se A si predica di ogni B tutti i B hanno la proprietà A, ecc.) rende trasparente l'implicazione. Tuttavia si è qui ben lungi dall'essere in un contesto matematico o afferente in qualche modo alla matematica. Occorrerà un processo di molti secoli, fino all'Ottocento, perché si realizzi quell'unità fra logica e matematica che ha prodotto la «logica matematica» moderna.

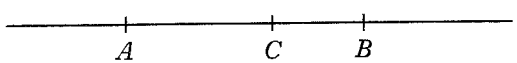
Si può ammettere, certo, che gli *Elementi* abbiano un'organizzazione conforme alla logica aristotelica, ma ciò è vero in un senso molto generale, per quanto riguarda l'architettura complessiva dell'opera e non nel senso che la logica vi compaia come uno «strumento» per sviluppare singole dimostrazioni od organizzare complessi di dimostrazioni. Vi sono negli *Elementi* forme deduttive non sillogistiche e, in generale, le dimostrazioni sono modellate sull'oggetto matematico piuttosto che ricondotte a tipi standard dal punto di vista logico.

È molto importante sottolineare questa autonomia del discorso matematico, autonomia che diverrà ancora piú marcata con il sorgere della scienza nuova, per ben capire come l'unità che si viene a realizzare tra matematica e logica nel secolo scorso presuppone un determinato grado di sviluppo delle due discipline. L'esperienza dell'algebra è, ovviamente, fondamentale nell'opera di Boole, ma non è solo l'«algebrizzazione della logica» che segna la differenza con Aristotele, è l'intero modo di intendere la disciplina in connessione con l'attività scientifica.

Si tratta di osservazioni molto semplici. Esse suggeriscono tuttavia come spesso anche a una lettura «accurata» dei testi e a una interpretazione fondamentalmente «esatta», come in questo caso della breve frase citata di Aristotele, possono corrispondere non poche drastiche semplificazioni.

Anche l'opposizione ♦continuo/discreto♦ che inerisce ad aspetti fondamentali dell'esperienza umana, ha suscitato, nell'antichità, le riflessioni teoriche piú profonde nel pensiero greco. Tra le piú celebri vi sono i paradossi di Zenone sul moto, esposti e analizzati da Aristotele; in essi, forse piú che altrove, si avverte la profonda tensione concettuale che pervade l'opposizione. Una «soluzione» dei paradossi di Zenone che faccia capo a una interpretazione univoca e corrisponda a un'unica idea risolutiva non si è data fino ad ora, né può darsi se nella coppia continuo/discreto si manifesta una contraddizione effettiva. Ma il problema certo può essere approfondito e la formulazione moderna della continuità della retta data da Dedekind costituisce un progresso fondamentale. Già Aristotele [*Fisica*, 233a, 24-28] aveva osservato come «in due sensi si dicono infiniti tanto la lunghezza quanto il tempo e in genere ogni continuo: o per divisione o per gli estremi. Pertanto, gli infiniti che sono tali secondo la quantità, non possono toccarsi in un tempo finito; quelli, invece, che sono tali secondo la divisione, lo possono perché il tempo stesso è infinito sotto questo aspetto» (cfr. ♦Divisibilità♦).

Il continuo è infinito per divisione. Questo aspetto della continuità è colto con grande chiarezza da Aristotele. Se si considera il caso della retta, ciò significa che tra due suoi punti A e B c'è sempre un C intermedio:



Ma questo non è il solo aspetto e ancora Aristotele lo avverte con grande chiarezza, tant'è che l'aporia di Zenone non è ricondotta a questo solo aspetto ma all'essere il tempo e lo spazio, sotto questo rispetto, uguali. Questo è un modo certo geniale per rendere pensabile il moto, ma lascia inesplorato un aspetto della continuità d'un ente preso per se stesso isolatamente.

Dedekind, con uno scritto del 1872 (l'anno stesso del Programma di Erlangen) reca un contributo decisivo al problema della comprensione della continuità. Si tratta del saggio *Continuità e numeri irrazionali* (*Stetigkeit und irrationale Zahlen*). Vi si osserva come, considerati i numeri razionali con la loro relazione d'ordine, un numero a ripartisca tutti i numeri razionali in due classi A_1 e A_2, entrambe infinite, delle quali la prima contiene tutti gli a_1 con $a_1 < a$ e la seconda tutti gli a_2 con $a_2 > a$ (a stesso può porsi nella prima o nella seconda classe).

Questa proprietà trova riscontro in una consimile proprietà dei punti di una retta: un punto A divide i punti della retta in due classi, formata la prima dai punti A_1 alla sinistra di A e la seconda dai punti A_2 alla destra di A (analoga osservazione vale per A). Cosa ora fa sí che la retta sia continua e che tale non sia il sistema dei numeri razionali?

Se si fissa un riferimento sulla retta, è possibile disporvi al modo solito tutti i numeri razionali e, come è ben noto, la presenza dei segmenti incommensurabili si manifesta nel fatto che vi sono punti sulla retta che non corrispondono a numeri razionali (come il segmento che rappresenta la diagonale del quadrato che ha per lato l'unità di misura).

Ebbene questa differenza, questa «mancanza» di numeri corrisponde esattamente al fatto che ogni suddivisione dei punti della retta in due classi tali che ogni punto della prima classe stia a sinistra di ogni punto della seconda è realizzata esattamente da un punto, mentre ciò non avviene per le classi di razionali tali che ogni numero della prima sia inferiore a ogni numero della seconda. Non vi sono partizioni dei punti della retta che abbiano rispetto all'ordinamento la stessa caratteristica di quelle che provengono dai suoi punti che *già non provengano da punti*.

«Che ognuno trovi il principio enunciato tanto evidente e tanto concordante con la sua propria rappresentazione della retta, – osserva Dedekind, – ciò mi soddisfa al massimo grado, perché né a me né ad altri è possibile dare di questo principio una dimostrazione qualsiasi. La proprietà della retta espressa da questo principio non è che un assioma, ed è solo sotto forma di questo assioma che noi pensiamo la continuità della retta, che riconosciamo alla retta la sua continuità».

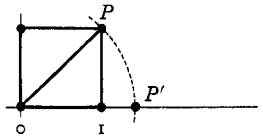


Figura 2.

Il punto P' sulla retta non è rappresentato da alcun numero razionale (corrisponde infatti a $\sqrt{2}$).

Con la formulazione di Dedekind della continuità si ha una concezione della retta che è molto piú ricca e articolata di quella presente nella matematica classica: il continuo non è tale solo per la ◆divisibilità◆ infinita (cosa, come si è visto, già osservata da Aristotele) e un'analisi del moto che si fondi solo su questa proprietà non è adeguata. Ciò non risolve l'aporia di Zenone ma la rende «piú pensabile» poiché non ci si limita ad affermare che una certa connotazione non esaurisce il concetto della continuità, ma si dà di quest'ultima un'ulteriore caratteristica: la retta come ◆insieme◆ ordinato è analizzata piú profondamente.

Tuttavia, come è ben noto, Dedekind, dopo aver «creato» i numeri reali in modo che il continuo numerico abbia le stesse proprietà di continuità della retta, rifiuta a quest'ultima (in un approccio fondazionale, non certo nella didattica!) un'uguale dignità. Giunge cosí a compimento un desiderio che fa capo a una tradizione antichissima: l'intera matematica ridotta ai numeri. Al tempo stesso la geometria se non sacrificata, come scienza autonoma, certo viene diminuita.

L'esempio piú celebre di ◆dualità◆ è quello dato dal piano proiettivo. L'osservazione che i punti e le rette in un piano ordinario hanno comportamenti «analoghi» è certo molto antica. Due punti arbitrari e distinti individuano una retta, cosí come due rette in generale, cioè se non sono parallele, individuano un punto. La restrizione «non parallele» è stata di fatto eliminata nella pratica del disegno (ove ogni costruzione che fa intervenire una coppia di rette parallele – una direzione – si svolge in modo del tutto simile alla costruzione corrispondente ove le rette non siano parallele) ben prima che avesse un fondamento teorico la nozione di «punto all'infinito». All'inizio del secolo scorso, aggiunti al piano ordinario i «punti all'infinito» ordinatamente disposti sulla «retta all'infinito», si hanno le premesse per il piano proiettivo, l'ambiente naturale per la dualità fra punti e rette.

Su due fatti si vuol richiamare l'attenzione, due fatti che non costituiscono un semplice prolungamento del vecchio nel nuovo: i punti all'infinito permettono di non far distinzione in una costruzione grafica tra elementi al finito e all'infinito, unificano costruzioni all'apparenza assai diverse. È la stessa cosa assegnare un'iperbole mediante i suoi asintoti e un suo punto oppure mediante due tangenti con il loro punto di contatto e un suo punto. La costruzione di un arbitrario punto ulteriore procede in modo del tutto conforme, per fare un esempio. Ma non è per questo e per altre consimili costruzioni che si dispone del piano proiettivo. A questo punto si dispone soltanto del piano ordinario con l'aggiunta di elementi ideali; si ha il piano proiettivo quando i nuovi enti appaiono come del tutto *identici* ai punti ordinari, quando non c'è piú la retta impropria ma ogni retta ha lo stesso ruolo. Occorre un poderoso sforzo di astrazione per prescindere dalle qualità all'apparenza piú concrete di punti e rette e concepirli come definiti dal loro reciproco interagire (un inizio, come si vede, del discorso di Hilbert).

Ancora, perché abbia senso la dualità occorre una distinzione fondamentale: se un linguaggio, con una semplice sostituzione terminologica, descrive due realtà diverse, è ovvio che *lo stesso* linguaggio (i nomi dei termini sono arbitrari) può descrivere differenti realtà. Si è perciò costretti a distinguere tra il linguaggio e ciò che il linguaggio descrive. È un esempio fondamentale del concetto di modello. Come si vede, anche se il concetto di dualità sembra legarsi a fatti tra i piú semplici, esso diviene disponibile solo dopo acquisizioni teoriche di grande rilievo.

Negli esempi esaminati sino ad ora, il confronto tra tematiche classiche e idee innovative era in qualche modo contenuto dalla natura dell'argomento. Con il gruppo Bourbaki, autore di un'opera tuttora in esecuzione, dal titolo assai significativo di *Eléments de mathématique* (1939 sgg.), si pone coscientemente l'idea di una matematica globalmente «nuova».

Bourbaki rinunzia a una «filosofia» della matematica (rinunzia spesso enfatizzata da qualche componente del gruppo) proponendosi piú «modestamente» di catalogare e sistematizzare i metodi della matematica. Al tempo stesso Bourbaki sottolinea le possibilità euristiche del metodo assiomatico, purché esso sia inteso nella sua effettiva essenza, come strumento per aumentare le possibilità dimostrative.

Un ruolo essenziale nel progetto bourbakiano ha il concetto di ◆struttura matematica◆: l'organizzazione generale del discorso matematico secondo strutture algebriche, strutture d'ordine, strutture topologiche, ove il singolo oggetto matematico non è un «dato», ma il frutto di una molteplicità di determinazioni.

L'idea nuova e profonda della matematica come una totalità, come un'architettura ove i singoli oggetti acquistano la loro individualità solo in rapporto al tutto, è certamente un grande contributo di Bourbaki, ma l'idea della matematica come una totalità organizzata non è una caratteristica inscindibile dall'approccio organizzato attraverso il concetto di struttura matematica. Essa è presente anche nell'approccio categoriale.

Anche per questa ragione, la distanza che separa la descrizione categoriale degli oggetti matematici dalla loro descrizione strutturale non è forse cosí grande come spesso si crede. Giustamente Manin parla di una descrizione del primo tipo come «duale» rispetto a quella del secondo.

Dal punto di vista della teoria delle strutture, osserva Manin, un gruppo è *un ◆insieme◆ dotato di una determinata struttura*, mentre dal punto di vista della teoria delle categorie un gruppo *è un oggetto della categoria di tutti i gruppi*. Da questo secondo punto di vista un oggetto riceve la propria identità dal coesistere con altri oggetti della stessa natura, dal modo con il quale esso è posto in relazione con loro. Ma anche dal punto di vista bourbakiano il concetto di gruppo appare come determinato da una pluralità di confronti, confronti a livello di struttura, confronti che conducono alla scelta di quella particolare struttura che è tale non in sé ma in confronto ad altre strutture ad essa paragonabili. È ovvio che l'affermare che una certa legge di composizione è associativa non significa solamente disporre di una «proprietà» per la legge in questione ma anche, e forse principalmente, porre una differenza specifica con altre leggi di composizione che non sono associative.

L'approccio categoriale sembra invece proporre aspetti differenti e forse ancor piú innovativi per due riguardi: per quanto riguarda il «divenire» e per quanto riguarda il concetto di «fondamentale».

«La dialettica della opposizione struttura/evoluzione, – osserva Manin, – ri-

flette la generale opposizione scientifica essere/divenire. Il successo della teoria degli insiemi come linguaggio universale della matematica e della concezione delle strutture matematiche come realizzazione concreta di questo linguaggio ha segnato uno spostamento di accento sui primi elementi di queste opposizioni. Le concezioni relative sempre piú diffuse nei fondamenti, gli assiomi categoriali, ecc. segnano il movimento inverso, quando "divenire" e "incompiuto" sono considerati primitivi».

In molti articoli di questa *Enciclopedia* si è accennato all'approccio categoriale come recante in sé questo significato di apertura verso il «divenire» e non è il caso di ripercorrere gli esempi. Vale la pena invece di riflettere sull'importanza di un'assunzione teorico-programmatica di questo tipo. Forse è presto per trarre conclusioni significative, ma ad una almeno sembra si possa accedere: accanto all'approccio categoriale è necessario valutare attentamente la storia della matematica degli ultimi trecento anni.

È chiaro infatti che se l'aritmetizzazione dell'analisi e la riduzione della matematica alla teoria degli insiemi appaiono come processi del tutto necessari, se appare necessaria l'introduzione del «rigore» ottocentesco per eliminare ogni appello all'intuizione del continuo, l'approccio categoriale, privato di una legittimità storica, può apparire al piú un utile «complemento» all'approccio insiemistico, ma non può pretendere a una pari importanza. Se invece le acquisizioni delle quali s'è detto appaiono «storicamente» date e frutto di scelte determinate, ecco che nell'esame di queste scelte è possibile costruire un percorso che mostri come anche l'approccio categoriale può giudicarsi erede di un processo concreto.

Ora è significativo a questo riguardo come numerosi articoli recenti (di Lawvere, ma anche di Kock, Reyes, ecc.) mostrino la possibilità di affrontare molti problemi secondo tecniche che, sia pure in contesti molto sofisticati, ripropongono certi interventi «diretti» dell'analisi settecentesca, con un uso raffinato degli «infinitesimi».

Fin troppo ovvio dunque, che questo debba essere collegato a un esame accurato delle ragioni per le quali gli infinitesimi, e le nozioni correlate, sono stati eliminati dall'analisi.

Non si tratta naturalmente di «demolire» o sostituire la matematica classica. È ben possibile però (e non mancano esempi nella storia della matematica) che l'abbandono di determinati metodi abbia avuto come motivazione precipua la difficoltà d'una loro spiegazione teorica, la quale può ben essere poi possibile come frutto di acquisizioni datesi successivamente.

Il secondo aspetto innovativo è invece nel modo stesso di concepire i fondamenti. Esso è legato alla teoria delle categorie ma, se pure è eccessivo giudicare contingente questo legame, non è certo la teoria delle categorie a essere realmente in gioco.

Si è osservato che in molti tra i piú vari settori della matematica molte costruzioni fondamentali appaiono legate a categorie, funtori o trasformazioni naturali (cfr. l'articolo ◆Trasformazioni naturali / categorie◆): facilmente e chiaramente descrivibili in questi termini. In particolare, ha osservato Lawvere, il concetto di «coppia di funtori aggiunti» sembra il piú diffuso nella «pratica» matematica. Esso dunque, che appare naturalmente al centro del nostro agire concreto in quanto matematici, deve avere un pari ruolo come oggetto fondamentale della riflessione teorica. Esprimendosi con una certa semplificazione: fondare significa porre al centro della teoria ciò che già è al centro nella pratica matematica. Non la ricerca dell'atto piú semplice del nostro pensiero, o delle leggi elementari del pensiero, o dei primi fatti psicologici legati al pensiero matematico viene giudicato come fondamentale, ma ciò che l'analisi concreta mostra essere al centro della matematica stessa.

Si potrebbe naturalmente proporre una lunga lista di obiezioni a questo approccio (chi giudica cosa è al centro e perché? cosa vuol dire «pratica matematica»?, ecc.) e molte ne sono state poste. Ma preme piú osservare come ad esso si collegano molte questioni di grande rilievo. È ovvio intanto che questo concetto di fondazione è dato *storicamente*. Oggi il concetto di coppia di funtori aggiunti ha un ruolo fondamentale, ma per nessuna ragione, se non ponendo dei vincoli soggettivi agli sviluppi futuri, si può giudicare che cosí sarà anche in futuro. Ecco che allora la ricerca di ciò che è centrale si lega all'analisi del processo storico come esso concretamente si svolge sino al tempo della ricerca: la storia della matematica appare come parte del concetto stesso di fondazione (pur essendo, è ovvio, distinguibile una parte «operativa» da una parte storica).

Inoltre, ed è forse l'aspetto che Lawvere tende a sottolineare maggiormente, questo concetto di fondazione non è una riflessione su ciò che è la matematica ma vuole essere elemento essenziale per lo sviluppo della matematica stessa.

Ma a giudizio di chi scrive pare importante, anche se certo di minore importanza, che in questa prospettiva la storia della matematica non venga ad essere «accanto» alla matematica, ma organicamente inserita in essa. [M. G.].

Bourbaki, N.

1948 *L'architecture des mathématiques*, in F. Le Lionnais (a cura di), *Les grands courants de la pensée mathématique*, Cahiers du Sud, Marseille; nuova ed. Blanchard, Paris 1962, pp. 35-47.

1954 *Eléments de mathématique*, I. *Théorie des ensembles* (fasc. 17), Hermann, Paris.

Browder, F. E.

[1974] (a cura di) *Mathematical Developments Arising from Hilbert Problems. Symposium in Pure Mathematics, Northern Illinois University, 1974*, American Mathematical Society, Providence R.I. 1976.

Gödel, K.

1931 *Über formale unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, in «Monatshefte für Mathematik und Physik», XXXVII, pp. 349-60 (trad. it. in E. Agazzi, *Introduzione ai problemi dell'assiomatica*, Vita e Pensiero, Milano 1961, pp. 203-28).

Grünbaum, A.

1967 *Modern Science and Zeno's Paradoxes*, Wesleyan University Press, Middletown Conn.

Halmos, P. R.

1957 *Nicolas Bourbaki*, in «Scientific American», XCVI, 5, pp. 88-99.

Lawvere, F. W.

[1967] *Categorical Dynamics*, in A. Kock (a cura di), *Topos-Theoretic Methods in Geometry and Analysis*, Matematisk Institut, Aarhus Universitet, Aarhus 1978.

1969 *Adjointness in foundations*, in «Dialectica», XXIII, 3-4, pp. 281-95.

MacLane, S.
1971 *Categories for the Working Mathematician*, Springer, Berlin - Heidelberg - New York (trad. it. Boringhieri, Torino 1977).

Neumann, J. von
1958 *The Computer and the Brain*, Yale University Press, New Haven Conn. (trad. it. in V. Somenzi (a cura di), *La filosofia degli automi*, Boringhieri, Torino 1965, pp. 157-221).

Thom, R.
1977 *Stabilité structurelle et morphogénèse. Essai d'une théorie générale des modèles*, Interéditions, Paris (trad. it. Einaudi, Torino 1980).
1981 *Modèles mathématiques de la morphogénèse*, Bourgois, Paris.

Weyl, H.
1952 *Symmetry*, Princeton University Press, Princeton N.J. (trad. it. Feltrinelli, Milano 1962).

Whitehead, A. N., e Russell, B.
1962 *Principia Mathematica to *56*, Cambridge University Press, London.

Applicazioni

1. *Nozione generale ed esempi.*

La nozione di applicazione è un modello matematico molto generale di aspetti vari della realtà e del pensiero. Essa si genera quale forma dell'idea di una *corrispondenza a coppie di oggetti di una classe, o di un insieme, con oggetti di un'altra classe.* Una tale corrispondenza si realizza nella descrizione di processi che si sviluppano nel tempo, quando ad ogni momento del tempo corrisponde un certo stato del sistema studiato. Essa viene usata nello studio di operazioni su degli oggetti: il risultato dell'operazione viene associato all'oggetto sul quale l'operazione viene eseguita. Essa può servire per descrivere dipendenze di genere diverso nei casi in cui una certa caratteristica di una situazione è completamente definita da un'altra caratteristica. Essa chiarisce l'immagine che si ha di simmetria, di omogeneità, ecc.; nella simmetria di un ornamento, o di un cristallo, si tratta di applicazioni dei suoi elementi (dei particolari del rabesco, degli atomi), le quali conservano la struttura dell'ornamento, o del reticolo del cristallo.

Nell'ambito della teoria degli insiemi la nozione di applicazione è secondaria rispetto a quella di insieme. Si chiama applicazione, oppure (in un uso piú restrittivo) funzione di un insieme X in un insieme Y una corrispondenza f, la quale ad ogni elemento x di X associa un unico elemento y di Y, spesso indicato con $f(x)$. Qui x si chiama argomento dell'applicazione f, mentre $f(x)$ si dice il suo valore sull'argomento (o nel punto) x. L'insieme X si chiama dominio di definizione (o semplicemente dominio) di f, mentre l'insieme Y si chiama dominio dei valori (o codominio) di f. L'applicazione f con il dominio X e il codominio Y spesso si scrive cosí: $f : X \to Y$. Negli impieghi pratici il caso piú frequente è quello in cui X e Y sono insiemi di numeri; interi, razionali, reali, oppure complessi, che possono essere considerati quali risultati di possibili misurazioni di queste o quelle caratteristiche della realtà. Ma nella matematica teorica si studiano applicazioni d'insiemi assai piú generici. Se l'insieme X è finito, l'applicazione f può essere data con una tabella, nella quale accanto ad ogni possibile valore dell'argomento $x \in X$ si scriva il valore $f(x)$. Esempi ben noti sono le tavole dei logaritmi, delle funzioni trigonometriche e di altre funzioni speciali. (A rigore, queste tavole dànno non la stessa funzione f, ma solo una parte dei suoi valori su una parte dei valori dell'argomento e, per di piú, approssimativamente, a meno di un certo numero di cifre decimali dopo la virgola). Un altro modo di dare una funzione numerica consiste nell'esibire il suo grafico. Infine un'applicazione può essere data per mezzo della descrizione della procedura che, applicata ad ogni elemento $x \in X$ del dominio di f e dopo la sua esecuzione, fornisce il valore $f(x)$. Hanno questa caratteristica le regole scolastiche per calcolare la somma, o il prodotto di due numeri dati nella scrittura decimale, e, in generale, la maggioranza delle «formule» per le funzioni che per-

mettono di calcolare i loro valori. Se la descrizione della procedura per il calcolo è cosí completa che la sua esecuzione può essere affidata a un dispositivo meccanico, per esempio a un calcolatore, essa si chiama *algoritmo*, mentre la corrispondente applicazione f si dice calcolabile algoritmicamente.

Quella che segue è la descrizione particolareggiata di alcuni esempi di applicazioni e dei meccanismi della loro realizzazione nel modellare situazioni fisiche, umane e sociali.

1.1. Stato di un sistema fisico quale funzione del tempo.

Uno dei processi fisici piú semplici è il moto di un corpo nello spazio. Spesso si può fare astrazione dalle dimensioni finite del corpo e considerarlo quale punto materiale, cosí che la sua posizione può essere data dallo spazio da esso percorso lungo la sua traiettoria. Si può immaginare il moto di una pietra che cade, di una automobile, o di un satellite della Terra. Supponiamo che lo spazio e il tempo siano misurati in prescelte unità di misura e che sia fissato il momento iniziale del tempo. Allora ad ogni momento del tempo corrisponde lo spazio percorso $r(t)$. Se si fa astrazione dalla limitatezza del tempo di osservazione, si può ritenere che t assuma tutti i valori reali ≥ 0 («futuro»), oppure persino tutti i valori reali («futuro» e «passato»), e r sarà una funzione R(numeri reali = tempo) $\to R$(numeri reali = spazio). In un modello piú realistico la funzione r è definita solo in un certo intervallo finito; ciò, spesso, va tenuto presente anche quando si descrivono sistemi a lunga vita, quali il sistema solare, o l'universo.

Una descrizione completa della posizione di un punto materiale è data da tre coordinate: per esempio, dalle distanze da tre piani fissi perpendicolari tra loro. Perciò il moto del punto è completamente descritto dalle variazioni di queste coordinate quali funzioni del tempo, cioè da una applicazione R(tempo) $\to R^3$ (spazio), dove R^3 è l'insieme delle terne ordinate di numeri reali, cioè il prodotto cartesiano $R \times R \times R$.

Piú in generale, sia S un certo sistema fisico, il cui stato, in ogni momento del tempo, sia dato da certe «coordinate generalizzate» reali, $q_1, \ldots, q_n$. Per esempio, se S è il sistema solare, $q_1, \ldots, q_n$ possono essere le coordinate (nel sistema connesso con il Sole) di tutti i pianeti che, in una prima approssimazione, possono essere considerati quali punti materiali. In questo caso l'evoluzione del sistema è data da una applicazione R(tempo) $\to R^n$(spazio delle coordinate generalizzate, o spazio configurazionale del sistema), cioè da un insieme di n funzioni, $q_1(t), \ldots, q_n(t)$. Un altro esempio è dato dal gas in un cilindro con stantuffo; per esempio, dal vapore in un cilindro di una macchina a vapore. Il suo stato, in ogni momento, è definito da tre numeri, V, p, T (volume, pressione e temperatura), mentre le funzioni $V(t)$, $p(t)$, $T(t)$, che descrivono la dipendenza di queste caratteristiche dal tempo, contengono l'informazione fondamentale sul lavoro di una tale «macchina termica». L'insieme di tutti i valori ammissibili delle coordinate generalizzate, $q_1, \ldots, q_n$, non coincide necessariamente con tutto lo spazio R^n: il moto di un sistema può essere limitato da vincoli che individuano solo una parte U di questo spazio. Per esempio, la posizione di un segmento materiale, del genere di un righello (se si trascurano la sua larghezza e spessore, ma non la lunghezza), è definita dalle coordinate cartesiane dei suoi estremi, diciamo (q_1, q_2, q_3) e (q_1', q_2', q_3'); però esse non possono essere arbitrarie in R^6, ma devono essere soggette alla condizione di conservare la lunghezza l del righello: $(q_1-q_1')^2+(q_2-q_2')^2+(q_3-q_3')^2=l^2$, e questa condizione individua un sottoinsieme $U \subset R^6$. Il moto di un tale segmento è dato da una applicazione $R \to U$.

Nella teoria dei mezzi continui, quali i liquidi e i gas, i campi elettrici ed altri, lo stato di un sistema in ogni momento del tempo è dato da un insieme di numeri non finito, ma infinito, che è ugualmente comodo considerare quale applicazione (cfr. oltre).

Il tempo è l'argomento universale delle funzioni che descrivono tutti i possibili processi fisici, astronomici, geologici, storici, economici, tecnici e di altro genere. Storicamente la comprensione di una tale possibile descrizione dei processi ha avuto una enorme importanza nello sviluppo della matematica e delle scienze naturali. Probabilmente l'idea fondamentale fu quella secondo la quale in una tale descrizione si potevano ritenere preassegnati «contemporaneamente», «staticamente», tutti i momenti del passato e del futuro, benché il passato scompaia continuamente e il futuro appaia sotto forma di un istante fuggevole e non sia mai dato come realtà. Solo questa idea permette di sperare nell'adeguatezza della descrizione di un processo mediante una applicazione $R \to U$, dove la retta dei tempi R intervenga quale oggetto matematico unico. A sua volta, la formazione di questa idea si è basata sull'osservazione di processi dotati di una netta regolarità e importanti per lo sviluppo dell'economia e della società, quali il moto delle stelle (orientamento nella navigazione) e gli straripamenti dei grandi fiumi (determinazione del tempo delle semine). Le tabelle e i metodi per calcolare il moto delle stelle, le eclissi solari, ecc., che risalgono all'epoca in cui ebbe origine la civiltà, sono l'incarnazione di questa idea.

L'efficacia del considerare l'insieme dei momenti del tempo come un tutto unico ha predisposto all'accettazione del punto di vista generale teorico-insiemistico, dove di solito il dominio X di un'applicazione modella l'insieme delle «diverse possibilità», una parte o la maggioranza delle quali non si realizza mai nel sistema dato. In questioni di contabilità, o in calcoli tecnologici, probabilmente non s'incontreranno mai numeri maggiori, diciamo, di 10^{1000}; tuttavia, l'algoritmo per la somma è formulato, e a buon diritto, per tutti i numeri interi: solo cosí l'aritmetica diventa effettivamente utile e comoda nell'uso.

Le idee connesse con lo sviluppo dei processi nel tempo furono alla base di molte nozioni matematiche piú specifiche. Per esempio, il problema di calcolare la velocità istantanea e l'accelerazione di un punto materiale in un moto arbitrario fu risolto da Newton con l'invenzione del calcolo differenziale. Fra gli esempi che si sono avuti piú tardi vi è il cosiddetto «problema di Cauchy» per diverse classi di equazioni differenziali, dove la premessa è l'idea secondo la quale è possibile predire univocamente l'evoluzione di un sistema, se ne sono noti lo stato e la velocità della sua variazione al momento iniziale del tempo.

Ancora, l'immagine fisica del moto di un punto materiale lungo una determinata linea – la sua traiettoria – è strettamente legata alla nascita della geometria analitica nelle opere di Descartes [1637]. Questi iniziò lo studio sistematico dei grafici delle funzioni numeriche f, quali sottoinsiemi del piano $R \times R$, costituiti da tutti i possibili punti con le coordinate $(x, f(x))$ e, piú in generale, dei grafici delle dipendenze funzionali «implicite» del tipo $f(x, y) = 0$, del genere dell'equazione della circonferenza $x^2+y^2-1=0$. Lo sviluppo di questa idea in senso concettuale ha condotto all'attuale identificazione di una funzione $f: X \to Y$ con il suo grafico nel prodotto cartesiano $X \times Y$, quale sottoinsieme di tipo determinato. Questa espressione della nozione di funzione nel linguaggio della teoria degli insiemi sarà esaminata nei particolari piú oltre.

Per concludere, è da notare ancora un altro aspetto importante della descrizione matematica dei processi evolutivi direttamente legato alla possibilità di considerare la retta dei tempi come un oggetto unico. Le leggi dell'evoluzione si possono dividere grosso modo in due classi: locali e globali. Un esempio di legge locale dell'evoluzione è la regola di Newton che descrive il moto di un punto materiale sotto l'azione di una forza: in ogni momento del tempo l'accelerazione del punto è proporzionale alla forza che su esso agisce. Suddividendo l'intervallo di tempo durante il quale si segue il moto del punto in tante piccole parti, è possibile calcolare approssimativamente questo moto, ritenendo che la velocità resti costante in ognuna di queste parti, e sia quindi accresciuta a salti della quantità prescritta dalla legge di Newton. Il passaggio al limite conseguente al rimpicciolimento infinito della suddivisione significa *integrare* questa legge locale del moto. Piú tardi si notò che il moto nel suo insieme può essere descritto globalmente mediante il cosiddetto principio variazionale: il moto di un punto viene definito minimizzando sulla sua traiettoria una certa grandezza fisica che si chiama azione. Questo principio presuppone la definizione preliminare di azione quale funzione su tutte le possibili traiettorie, persino quelle fisicamente irrealizzabili. Negli ultimi decenni questo principio ha avuto un profondo sviluppo nella trattazione della meccanica quantistica secondo Feynman, dove l'indefinibilità della traiettoria di una microparticella permette di attribuire un certo significato all'idea che la particella compia il moto «simultaneamente lungo tutte le traiettorie».

1.2. Campi.

I sistemi e i processi fisici quali il flusso di un liquido o di un gas, la distribuzione del pulviscolo di materia infrastellare, la propagazione del calore, gli effetti elettromagnetici e gravitazionali e molti altri sono descritti in termini di evoluzione dei campi. Il tipo piú semplice di campo – campo scalare in un dominio D dello spazio fisico – è modellato matematicamente da una applicazione della forma $D \to R$, oppure $D \to C$: ad ogni punto di D è associato un numero reale, o complesso. Questo numero può corrispondere alla densità media della materia in un piccolo dominio attorno al dato punto, alla temperatura nel dato punto, ecc. Il campo gravitazionale nella fisica classica è un esempio

di campo vettoriale: al punto è associato un vettore, la risultante di tutte le forze di attrazione gravitazionale in quel punto. Il flusso di un liquido in ogni momento del tempo pure è descritto da un campo vettoriale, il campo delle velocità del flusso. Lo stato istantaneo di un campo elettromagnetico nella teoria classica è descritto da due campi vettoriali nel dominio D: il campo delle tensioni dell'interazione elettrica e il campo delle tensioni dell'interazione magnetica. I valori del campo in un punto possono avere una natura ancora piú complessa; quella di tensori di dimensioni e di varianza diverse (tensore delle tensioni nella teoria dell'elasticità; tensore energia-impulso). Per la descrizione dei campi, quali applicazioni, esiste un apparato matematico speciale: la teoria delle varietà differenziabili e i fibrati su esse. Al dominio D corrisponde, per esempio, il suo fibrato tangente TD, costituito dai vettori tangenti in tutti i possibili punti di D; l'applicazione strutturale $p: TD \to D$ associa ad ogni vettore tangente la sua origine; il campo vettoriale corrisponde all'applicazione sezione $s: D \to TD$ con la proprietà $p(s(x)) = x$ per ogni $x \in D$.

La nozione di campo descrive bene un sistema fisico in quei domini dove non vi sono sorgenti di liquido, o di calore, accumulo di materia gravitante, cariche elettriche, ecc. Questi effetti corrispondono a singolarità del campo nel senso generale della parola: rottura della continuità, valori infinitamente grandi del campo, salti d'ogni genere e cosí via. Lo studio delle singolarità, della loro evoluzione nel tempo, della loro stabilità, ecc., fa parte dei problemi piú importanti relativi alla maggior parte dei modelli matematici dei processi continui. La corrispondenza «campo-singolarità» è un esempio tipico del dualismo generale «continuo-discreto».

Nella moderna teoria quantistica del campo si considerano campi di struttura ancora piú complessa. Nella meccanica quantistica classica ad ogni grandezza fisica (coordinata, impulso, energia, ecc.) viene associato un certo operatore lineare (cfr. § 1.5). Corrispondentemente il posto dei campi classici delle grandezze viene occupato dai campi di operatori. Questi oggetti matematici sono molto complessi e non esiste ancora una loro teoria conseguente, che ammetta un confronto diretto con la realtà fisica. Una delle difficoltà è connessa con le singolarità del campo che si hanno nelle sue origini, per esempio, negli elettroni per la teoria del campo elettromagnetico. È, questo, il cosiddetto problema delle divergenze della teoria del campo, che i fisici risolvono elaborando metodi di rinormalizzazione. Questi metodi forniscono un esempio notevole di frammenti di matematica efficaci, benché, a rigore, siano internamente contraddittori.

1.3. Traduzione e codificazione.

L'uso da parte dell'uomo di lingue naturali diverse, quale mezzo fondamentale per comunicare, registrare e conservare il sapere, conduce alla necessità di realizzare la traduzione dei testi da una lingua all'altra conservandone il significato. Modellare una traduzione con la nozione di applicazione non sarebbe completamente adeguato per due motivi. Primo, lo stesso insieme dei

testi che hanno significato nella lingua non è nettamente definito. Secondo, la traduzione notoriamente non è univoca persino al livello di singole parole: si veda un qualsiasi vocabolario bilingue. Assai meglio la nozione di applicazione è applicabile alle traduzioni da un linguaggio artificiale ad un altro. I linguaggi artificiali di programmazione, quali il Fortran, l'Algol e molti altri, furono elaborati quali mezzi di comunicazione fra l'uomo e i calcolatori. La loro struttura è approssimativamente uguale a quella di frammenti del gergo matematico umano, ciò che facilita il loro uso da parte dei programmatori, ma rende i testi in questo linguaggio inaccettabili per la macchina calcolatrice. Il linguaggio interno di una tale macchina, al contrario, è oltremodo frazionato e lontano dalla facoltà mentale umana standard perché possa essere efficacemente usato nella compilazione di un programma appena complesso. Perciò un programma scritto dall'uomo, diciamo in Algol, dev'essere tradotto in un programma in codici di macchina. Anche questo processo di traduzione è di solito affidato alla macchina calcolatrice, la quale lo attua usando un programma-codificatore scritto una volta per tutte. È comodo considerare la traduzione di un programma in Algol in un programma in codici di macchina quale calcolo di un'applicazione-codificazione definita sull'insieme di tutti i possibili programmi in Algol regolarmente compilati. Poiché la traduzione è affidata a un calcolatore, essa è una applicazione algoritmicamente calcolabile (cfr. sopra).

Negli ultimi anni idee analoghe sono andate diffondendosi piú largamente sia in connessione con la teoria e la pratica della traduzione automatica delle lingue naturali, sia in forza di esigenze interne della linguistica. In particolare si sta elaborando un linguaggio-mediatore semantico universale fra tutte le lingue naturali. Un tale linguaggio deve servire per la registrazione diretta dei «significati» dei testi naturali, indipendentemente dalle loro specifiche espressioni nelle diverse lingue naturali e dai vari modi sinonimici di una tale espressione in ogni singola lingua. Perciò una delle prime esigenze per questo linguaggio-mediatore consiste nel fatto che la traduzione di ogni testo naturale (come prima, di una loro classe finita) dev'essere univoca, cioè dev'essere ben modellata da una conveniente applicazione. La comprensione di questa esigenza ha già condotto a individuare nuove e interessanti categorie linguistiche; per esempio, le parole-operatori, le cui funzioni consistono nel far variare in modo «unidirezionale» il significato delle parole da esse dipendenti. Per esempio, le parole 'buono', 'forte', 'molto', 'velocemente', ecc., sono realizzazioni di uno stesso operatore: 'segno di intensificazione', che nel linguaggio-mediatore può essere una unità espressiva.

Un'altra esigenza consiste nel fatto che la traduzione nel linguaggio-mediatore di un testo naturale dev'essere un'applicazione algoritmicamente calcolabile, cosí come la traduzione dall'Algol in codici di macchina, o, quanto meno, prossima a questo ideale.

Ancor prima sono sorti modelli matematici di traduzione nella teoria della codificazione. Questa teoria risale a due fonti: alla cifratura di messaggi nella lingua naturale per ottenerne la segretezza (in questioni militari, nella corrispondenza diplomatica) e alla codificazione di messaggi per essere poi trasmessi

mediante questo o quel canale di comunicazione (codice Morse, codice marinaresco con le bandiere, codificazione dei risultati di misurazioni mediante successioni di impulsi in canali di comunicazione tecnici, ecc.). Qui l'univocità delle procedure di codificazione e di decodificazione, che permetta di considerarle applicazioni, è solitamente posta a fondamento della stessa struttura del codice. Possono far eccezione i cifrari con scelte casuali onde aumentare la segretezza; però, anche in questo caso, la decodificazione è solitamente univoca; altrimenti le scelte casuali potrebbero alterare l'informazione trasmessa. Il modo di ottenere una buona cifratura consiste nell'eliminare al massimo le regolarità e conformità presenti nel testo iniziale (frequenza di simboli, ripetizione di gruppi di simboli, ecc.). Lo studio dei problemi che in tal modo si presentano è stato intrapreso e portato molto avanti, in particolare dal noto matematico e ingegnere americano Shannon [1948].

Lo studio e la progettazione di apparecchiature elettroniche per la elaborazione delle informazioni (amplificatori, contatori, calcolatori...) ha condotto alla famosa immagine della «scatola nera». Si chiama cosí un oggetto avente un «ingresso» e un'«uscita» e tale che all'osservazione si prestano i segnali in uscita dovuti ai diversi segnali in ingresso, ma non la struttura interna di questo oggetto che attua l'elaborazione dei segnali. Pertanto una «scatola nera» realizza una certa applicazione {insieme dei segnali in ingresso} → {insieme dei segnali in uscita}, ma non è noto come essa la realizza. Tale è la situazione anche nello studio iniziale di progettazione di queste apparecchiature. Ciò fu di grande stimolo per far penetrare uno stile «funzionale» nella mentalità della pratica ingegneristica. Dopo la seconda guerra mondiale, con la comparsa della cibernetica, queste idee furono estese ai modelli fisiologici, dove sovente solo l'applicazione stessa si presta a uno studio sperimentale, mentre i meccanismi della sua realizzazione sono oggetto di una ricostruzione a partire da essa e, ove sia possibile, da dati complementari indiretti. Per esempio, nelle indagini elettroencefalografiche si sottopone chi è oggetto della sperimentazione a questa o quella situazione; cosí, gli si presentano degli stimoli visivi di questo o quel tipo, che costituiscono i segnali d'«ingresso». Si osservano cosí dei potenziali, rilevati con elettrodi sul cranio, oppure introdotti nel cervello. Questi potenziali sono considerati quali segnali d'«uscita», e il cervello quale scatola nera. In termini analoghi si può descrivere la maggioranza degli esperimenti tipo «stimolo-risposta». Una stabilità del carattere della reazione agli stimoli di un dato tipo interviene quale proprietà fondamentale dell'applicazione nel corrispondente modello matematico.

L'idea della «scatola nera» corrisponde alla separazione generale in matematica fra la nozione di applicazione e la nozione che esprime il modo di realizzarla: calcolo, tabella, formula, algoritmo (cfr. sopra). Questa separazione non è stata subito individuata e regolarizzata. Essa è strettamente legata alla formazione delle idee teoretico-insiemistiche ed è tuttora oggetto di discussioni semifilosofiche, nonché all'origine dei problemi riguardanti la possibilità di rappresentare le diverse classi di funzioni con questo o quel mezzo. Piú oltre si analizzeranno nei particolari tali questioni; per il momento importa descrivere

un arco di nozioni importanti legate alle applicazioni calcolabili algoritmicamente.

1.4. Algoritmi e calcolabilità.

Si sono già ricordati gli esempi piú noti di algoritmo: le regole per calcolare in scrittura decimale la somma e il prodotto di due numeri interi dati, pure in scrittura decimale. Che queste regole abbiano un carattere del tutto meccanico è dimostrato dal fatto che esistono da molto tempo dispositivi meccanici (o elettronici) per eseguire questi calcoli: le macchine calcolatrici. Un altro esempio già ricordato è il programma-codificatore che realizza un certo algoritmo per l'elaborazione dei testi in un linguaggio di programmazione in testi nel linguaggio dei codici di macchina.

Piú in generale, supponiamo di avere un insieme (potenzialmente infinito) di dati; per esempio, di testi, cioè successioni finite di segni di un prefissato alfabeto finito. Potrebbero essere le scritture decimali, o binarie dei numeri interi (negli alfabeti rispettivamente $\{0, 1, ..., 9\}$ e $\{0, 1\}$), oppure i programmi in Algol, ecc. *Un algoritmo è un insieme di regole per l'elaborazione di tali testi in testi nello stesso, o in un altro, alfabeto.* Un algoritmo dev'essere dato mediante un numero finito di procedure del tipo seguente: *a*) devono essere descritti i passaggi elementari dell'elaborazione dei testi; *b*) devono essere indicate le regole che in modo univoco definiscono la successione delle scelte dei passaggi elementari in dipendenza del testo iniziale e, possibilmente, del carattere di tutti i passaggi già compiuti al momento attuale; *c*) devono essere indicati i criteri per la fine della procedura dell'algoritmo. Se, dopo l'esecuzione di un numero finito di passaggi, l'algoritmo finisce, si dice ch'esso è *applicabile al testo iniziale* e che *lo elabora nel testo ottenuto dopo il passaggio finale.*

Secondo il significato di queste definizioni l'algoritmo per la somma di due numeri decimali è applicabile ad ogni coppia di tali numeri. L'algoritmo per calcolare la radice quadrata di un numero decimale positivo a è applicabile ad $a=4$ e ad $a=0{,}625$, ma non è applicabile ad $a=2$. Esso diventa applicabile a tutti i numeri decimali positivi se viene convertito nell'algoritmo per il calcolo approssimato, completandolo con una regola finale; diciamo: «arrestarsi dopo aver calcolato la sedicesima cifra decimale dopo la virgola».

Nella logica matematica è stata compiuta un'indagine esauriente della classe delle applicazioni assegnabili mediante algoritmo. L'inconveniente tecnico, legato al fatto che l'insieme sul quale è dato l'algoritmo non è univocamente definito, può essere superato mediante un metodo tecnico tipo numerazione di Gödel dei testi su un alfabeto finito. Ciò permette di ricondurre lo studio di ogni applicazione calcolabile algoritmicamente allo studio di applicazioni date su un sottoinsieme di numeri naturali e a valori nello stesso insieme di numeri naturali.

In quest'ordine d'idee il risultato fondamentale consiste nel fatto che è stata individuata la classe delle funzioni (parzialmente) ricorrenti. Ogni concreta funzione ricorrente si calcola mediante una procedura i cui passaggi elementari

sono scelti da un elenco fissato una volta per tutte. La cosiddetta tesi di Church afferma che *tutte le applicazioni algoritmicamente calcolabili sono funzioni ricorrenti.*

I grandi e moderni calcolatori elettronici digitali, se non si considerano i limiti materiali sul tempo di calcolo e sul volume della memoria, sono in grado di calcolare ogni funzione ricorrente e quindi, d'accordo con la tesi di Church, realizzano tecnicamente ogni algoritmo per l'elaborazione di testi. In questo senso i grossi calcolatori sono universali.

1.5. Misura in meccanica quantistica.

La misura di una caratteristica di un sistema fisico, quale la posizione, o la velocità di un punto materiale, la pressione di un gas, la luminosità di una sorgente luminosa, ecc., appartiene al numero di operazioni fondamentali che stabiliscono il legame fra la teoria e la realtà. Nella fisica classica si presuppone, per principio, che una misura può essere eseguita in modo tale da alterare poco quanto si vuole lo stato del sistema e che, per ogni procedura di misurazione, possono essere elaborati metodi che garantiscono una piccolezza arbitraria di questo disturbo. A lungo questo principio è stato considerato fondamentale per conclusioni gnoseologiche, quali quella dell'esistenza obiettiva delle proprietà studiate di un sistema.

Tuttavia, nello studio del micromondo il postulato sulla possibilità di tali misure è risultato del tutto infondato. È sostanzialmente impossibile chiarire questo punto rimanendo nell'ambito delle idee della fisica classica. Il fatto è che, per una descrizione adeguata dei microprocessi, bisogna ricorrere ad un nuovo linguaggio e, soprattutto, a una sua nuova interpretazione, diversa per principio. Questo linguaggio è fortemente matematicizzato; volendo può essere considerato come un frammento di una parte speciale della matematica (l'analisi funzionale). Peraltro, sono noti diversi argomenti intuitivi a favore della rinuncia al postulato classico sulla misura. Si consideri, per esempio, il problema di una misura esatta della posizione e della velocità di un elettrone nel vuoto, immaginandolo classicamente come un punto materiale. Per misurare le coordinate dell'elettrone bisogna «illuminarlo» e misurare l'irradiazione da esso riflessa. Qui l'esattezza della misura è limitata dalla lunghezza dell'onda luminosa. Per aumentarla bisogna diminuire la lunghezza di quest'onda. Però, l'«illuminazione» dell'elettrone consiste nel fatto che su esso si dissipa un fotone (un quanto di luce), la cui energia cresce al decrescere della lunghezza d'onda. Perciò nella dissipazione una parte dell'energia del fotone, trasmessa all'elettrone in modo imprevedibile, può diventare arbitrariamente grande, ciò che imprevedibilmente altera la velocità (l'impulso) dell'elettrone.

Lasciando questi argomenti intuitivi si passi ora a una piú esatta chiarificazione del modello matematico del processo di misurazione nella fisica quantistica a partire dallo stesso esempio dell'elettrone nel vuoto (lo schema è semplificato: non si tiene conto dello spin dell'elettrone e si ammette ch'esso si muova con una velocità trascurabile rispetto alla velocità della luce). Lo stato

istantaneo dell'elettrone è descritto, invece che dalla coppia (coordinata, velocità) della fisica classica, dalla cosiddetta ψ-funzione, che è una funzione a valori complessi, definita nei punti dello spazio fisico R^3 (o in un dominio dove può essere localizzato l'elettrone). Il quadrato del modulo di questa funzione in un punto, moltiplicato per il volume di una piccola sfera con il centro in questo punto, fornisce la probabilità di rinvenire l'elettrone in detta sfera. Finché non si effettua un esperimento reale per localizzare l'elettrone, la sua ψ-funzione varia nel tempo secondo la legge di evoluzione che solitamente viene descritta mediante l'equazione differenziale di Schrödinger. Se si vuole, si può immaginare che, invece del moto dell'elettrone, abbia luogo un moto dell'informazione su esso, espresso dalla ψ-funzione. Supponiamo che ora si faccia un esperimento per misurare le coordinate dell'elettrone. Secondo le idee ortodosse della meccanica quantistica, al momento della misura, la ψ-funzione dell'elettrone varia di un salto: se esso è stato rinvenuto in un piccolo dominio dello spazio, la sua nuova ψ-funzione sarà concentrata approssimativamente in questo dominio. Piú in generale, nel modello matematico della misura, ad ogni grandezza fisica (coordinata, energia, impulso, ecc.) è associato un operatore lineare, che è una applicazione dello spazio di tutte le possibili ψ-funzioni in se stesso. La sua linearità significa che, alla somma di due ψ-funzioni, esso associa la somma delle loro immagini, mentre al prodotto di una ψ-funzione per un numero associa il prodotto della sua immagine per lo stesso numero. Le predizioni probabilistiche circa i risultati della misura sono effettuate sulla base di speciali manipolazioni dell'algebra lineare applicate all'operatore di misura e alla ψ-funzione dell'elettrone al momento della misura. Questo operatore lineare è appunto il modello matematico fondamentale della misura.

1.6. Funzioni di utilità in economia.

In una serie di modelli matematici s'introducono oggetti, grandezze e applicazioni, la cui diretta osservazione è difficoltosa o, in generale, impossibile e, ciononostante, tali da condurre a una teoria che concorda con la realtà e la descrive bene. Le ψ-funzioni, viste nel punto precedente, sono probabilmente l'esempio piú chiaro. Convincono meno oggetti analoghi in modelli economici e sociali. Tuttavia, in virtú della loro notevole diffusione, si descriverà uno degli esempi piú semplici.

Esso si riferisce al cosiddetto modello di puro scambio. In questo modello si parte dall'idea di un insieme di merci di l tipi diversi e di un insieme di m consumatori $\{C_1, ..., C_m\}$. Se un dato consumatore ha x_i unità di merce del tipo i, il suo assortimento di merci è descritto dal punto $(x_1, ..., x_l) \in R^l$. Dal significato del modello segue che $x_i \geq 0$; inoltre il volume di merce per ogni tipo può ritenersi limitato, cosí che il vettore $(x_1, ..., x_l)$ giace solamente in un certo dominio P dell'ottante di dimensione l. Si presuppone che ogni consumatore C_i sia caratterizzato da una certa funzione di utilità $u_i : P \to R$, a valori reali. Il significato di questa funzione consiste nel fatto che C_i preferisce l'assortimento di merci $(x_1, ..., x_l)$ a quello $(x'_1, ..., x'_l)$ se, e solo se, $u_i(x_1, ..., x_l) > u_i(x'_1, ..., x'_l)$. Si presuppone che due consumatori possano scambiarsi una parte delle loro merci se, con questo, entrambe le funzioni di utilità crescono. In condizioni analoghe sono possibili anche scambi multipli.

Lo stato completo del modello è descritto da un punto di P^m, prodotto cartesiano m-multiplo del dominio P per se stesso, e questo punto è l'insieme degli assortimenti di merci di ognuno degli m consumatori. Uno stato si dice *ottimale secondo Pareto* se non vi è un altro stato nel quale tutte le funzioni di utilità non diminuiscono e almeno una cresce. L'idea consiste in questo: se lo stato del sistema non è ottimale secondo Pareto, può avvenire uno scambio vantaggioso. L'esistenza dello stato ottimale garantisce, per principio, la possibilità di raggiungere una situazione economica stabile.

1.7. Dipendenza.

I parametri che caratterizzano un sistema fisico, economico, o di altro genere, possono non essere indipendenti; i valori di alcuni di essi possono porre delle limitazioni ai valori degli altri. Se queste limitazioni sono talmente forti che il valore di uno di questi parametri è completamente definito dal valore degli altri, si ha una dipendenza funzionale. Il numero di esempi è enorme; moltissime leggi fisiche, particolarmente di «livello inferiore», sono formulazioni di tali dipendenze funzionali. Si è già ricordato che la temperatura T di un gas, per esempio, in un cilindro con uno stantuffo, può essere calcolata a partire dal suo volume V e dalla pressione p usando la formula $T = \text{cost}\ pV$. L'energia cinetica E di un punto materiale con massa m, che si muova con velocità v, è uguale a $(mv^2)/2$. Se sulla superficie di un corpo rigido viene mantenuta costante la distribuzione della temperatura, in ogni punto concreto di questo corpo si stabilisce una temperatura che dipende solo da questa distribuzione. Sono queste le fonti dell'insieme di funzioni, o applicazioni, che appaiono nei modelli matematici.

2. *Applicazione quale oggetto teorico-insiemistico.*

Come si è già ricordato, nell'ambito della teoria degli insiemi la nozione di applicazione non è primaria; essa si forma a partire da due nozioni fondamentali: «insieme» e «essere elemento». Il prodotto $X \times Y$ consiste di tutte le coppie ordinate $\langle x, y \rangle$, dove $x \in X$ e $y \in Y$. Nella teoria degli insiemi *si chiama applicazione dell'insieme X in Y un sottoinsieme $f \subseteq X \times Y$ tale che per ogni $x \in X$ esiste un unico elemento $y \in Y$ con la proprietà $\langle x, y \rangle \in f$.*

Sul carattere degli insiemi X, Y e del sottoinsieme $f \subseteq X \times Y$ nella definizione non si pone alcuna limitazione a priori. In particolare X e Y non sono necessariamente insiemi numerici; possono essere, a loro volta, insiemi di applicazioni, o di proprietà, oppure di queste, o di altre strutture matematiche, ecc. Il sottoinsieme f può essere dato con una «formula», oppure descritto mediante espressioni in un determinato linguaggio, ma ciò non è obbligatorio.

Tuttavia, la nozione di applicazione è cosí fondamentale, che una tale riduzione delle applicazioni agli insiemi spesso non viene attuata esplicitamente. In questo caso un'applicazione f viene considerata come un oggetto a sé, indipendente; al posto della notazione $\langle x, y\rangle \in f$ si usa la notazione $y=f(x)$, o una delle sue varianti ($y=fx$, $y=x^f$, ecc.); l'insieme delle coppie $\langle x, y\rangle \in f$ tali che $y=f(x)$ prende il nome speciale di *grafico* dell'applicazione f e può essere indicato con Γ_f, ecc.

Siano X, Y, Z tre insiemi e $f : X \to Y$, $g : Y \to Z$ siano applicazioni tra loro. A partire da esse si può costruire una terza applicazione detta composizione di f e g, $gf : X \to Z$, definita dalla formula $gf(x)=g(f(x))$ per ogni $x \in X$. La composizione di applicazioni è associativa: $h(gf)=(hg)f$, quando tutti i termini di questa uguaglianza sono definiti. Questa circostanza definisce in misura notevole il ruolo della proprietà associativa di diverse leggi di composizione in algebra. Ogni insieme X è dotato dell'applicazione canonica in se stesso identità: $\mathrm{id}_X : X \to X$, $\mathrm{id}_X(x)=x$ per ogni $x \in X$. Ogni applicazione composta con essa non cambia: per ogni $f : X \to Y$ si ha $f \circ \mathrm{id}_X = \mathrm{id}_Y \circ f = f$.

Nel linguaggio dell'analisi classica la composizione delle applicazioni si realizza spesso sostituendo all'argomento di una funzione i valori di un'altra funzione. Per esempio, $f(x)=\operatorname{sen} \log x$ è la composizione delle applicazioni $\log : \{\text{numeri reali positivi}\} \to R$ e $\operatorname{sen} : R \to R$. La composizione è uno strumento importante per costruire funzioni sempre piú complesse a partire dalle funzioni piú semplici. È in questo modo, per esempio, che si costruisce la classe delle funzioni elementari a partire dalle funzioni somma, prodotto, quoziente, esponenziale, radice, logaritmo e dalle funzioni trigonometriche, iterando la composizione in modi diversi. Esse sono molto importanti nelle applicazioni. Questa classe è sufficientemente estesa e comprende molte funzioni che con un buon grado di precisione modellano dipendenze e processi naturali e tecnologici. Come si è già ricordato, anche le funzioni calcolabili si ottengono per composizione a partire da un certo numero di semplici funzioni calcolabili (e con l'ulteriore processo speciale di ricorrenza). Nei calcoli automatici la composizione è spesso realizzata per mezzo di sottoprogrammi: per calcolare la funzione $g \circ f$ uno speciale sottoprogramma calcola i valori di f e questi sono poi dati in ingresso al sottoprogramma che calcola g.

Una applicazione $f : X \to Y$ si dice monomorfa, o iniettiva, se, per tutti gli x_1, x_2, da $x_1 \neq x_2$ segue che $f(x_1) \neq f(x_2)$; epimorfa, o surgettiva, se per ogni $y \in Y$ esiste $x \in X$ tale che $f(x)=y$; biettiva se è contemporaneamente iniettiva e surgettiva.

Per esempio, l'applicazione $f : R \to R$, $f(x)=x^2$, non è iniettiva poiché $f(-x)=f(x)$ per tutti gli x. Essa non è neppure surgettiva poiché per $y<0$ non esiste x tale che $x^2=y$. Restringendo i valori dell'argomento e della funzione ai soli numeri non negativi, essa diventa biettiva. Considerandola sui numeri complessi C l'applicazione $f(x)=x^2$ diventa surgettiva, ma non iniettiva. Se X è un insieme finito, l'applicazione $f : X \to X$ può essere iniettiva solo se è surgettiva e viceversa. Infatti qui ha senso parlare del numero dei diversi valori dell'argomento e della funzione e, per esempio, l'iniettività significa che il numero di valori diversi che assume f coincide con il numero di elementi di X e, perciò, per ogni $y \in Y$ si deve trovare un $x \in X$ tale che $f(x)=y$. È questo uno dei principî fondamentali della matematica combinatoria che studia gli insiemi finiti.

Si chiama immagine dell'applicazione $f : X \to Y$ il sottoinsieme $f(X)$ costituito da quegli $y \in Y$ per i quali esiste $x \in X$ con la proprietà $f(x)=y$. Si chiama controimmagine di un sottoinsieme $Z \subseteq Y$, e s'indica con $f^{-1}(Z)$, il sottoinsieme in X $\{x \mid f(x) \in Z\}$. Un'applicazione $f : X \to Y$ è surgettiva se l'immagine di f coincide con Y; è iniettiva se le controimmagini dei sottoinsiemi di Y con un solo elemento sono sottoinsiemi con un solo elemento, o vuoti, di X. In una proiezione p del piano cartesiano $R \times R$ su uno degli assi coordinati, come $p(x, y)=x$, la controimmagine di ogni punto x è tutta una retta. In cartografia, per rappresentare il rilievo della superficie terrestre sul piano della carta, si usano le linee di livello: queste linee sono le controimmagini dei diversi valori della funzione h : (regione terrena rappresentata) $\to$ (numeri reali) che dà l'altezza del punto dato sul livello del mare. Per analogia spesso si parla d'insiemi di livello di una data funzione, intendendo con ciò le controimmagini $f^{-1}(y)$ per i diversi $y \in Y$.

Ogni applicazione $f : X \to Y$ induce su X una relazione di equivalenza binaria $E_f \subseteq X \times X$, che consiste di tutte quelle coppie ordinate $\langle x_1, x_2\rangle \in X \times X$ per le quali $f(x_1)=f(x_2)$. Essa soddisfa i tre assiomi generali della relazione di equivalenza: $\langle x, x\rangle \in E_f$ per ogni x (riflessività); se $\langle x_1, x_2\rangle \in E_f$, anche $\langle x_2, x_1\rangle \in E_f$ (simmetricità); se $\langle x_1, x_2\rangle \in E_f$ e $\langle x_2, x_3\rangle \in E_f$, anche $\langle x_1, x_3\rangle \in E_f$ (transitività). S'indica con X/E_f l'insieme i cui elementi sono le classi di equivalenza di X rispetto a E_f, cioè i sottoinsiemi massimali di X costituiti dagli elementi equivalenti tra loro. È definita l'applicazione $s_f : X \to X/E_f$, dove $s_f(x)=$ classe di equivalenza contenente x, e cosí pure l'applicazione $i_f : X/E_f \to Y$, dove $i_f(\text{classe di } x)=f(x) \in Y$. L'applicazione s_f è surgettiva, mentre i_f è iniettiva; per ogni applicazione f si ha sempre la scomposizione $f=i_f s_f$ nella composizione di un'applicazione surgettiva e una iniettiva.

Le applicazioni canoniche sono legate a diverse costruzioni teorico-insiemistiche. Per esempio, il prodotto cartesiano $X \times Y$ è applicato surgettivamente sui suoi fattori mediante le proiezioni $p_X : X \times Y \to X$ e $p_Y : X \times Y \to Y$, dove $p_X(x, y)=x$ e $p_Y(x, y)=y$. Sia ora Z un insieme qualsiasi e consideriamo le due applicazioni $f_X : Z \to X$ e $f_Y : Z \to Y$. Esiste allora un'unica applicazione $f : Z \to X \times Y$ per la quale $f_X = p_X f$ e $f_Y = p_Y f$. È, questo, un esempio tipico della *universalità* della costruzione $(X \times Y; p_X, p_Y)$; si vedano gli esempi successivi.

Nella matematica classica si usano spesso delle generalizzazioni della nozione di applicazione; si tratta delle cosiddette applicazioni parziali e di quelle a piú valori.

Un'applicazione parziale dell'insieme X nell'insieme Y è un'applicazione $f : D \to Y$, dove $D \subseteq X$ è un certo sottoinsieme di X non necessariamente coincidente con tutto X. Per esempio, la funzione $f(x)=1/x$ da R in R non è definita, con questa formula, nel punto $x=0$ e, pertanto, è solo un'applicazione

parziale di R in R. Se f è un'applicazione parziale di X in Y, definita sul sottoinsieme $D \subseteq X$, si dice spesso che f ha delle *singolarità* nel complemento $X \setminus D$. Questa terminologia si assume solitamente quando $X \setminus D$ è piccolo rispetto a D, per esempio quando consiste di un numero finito di punti in un dominio numerico, e quando nelle approssimazioni $x \in D$ ai punti di $X \setminus D$ si rivelano delle irregolarità nel comportamento della funzione. In generale la nozione di singolarità di una funzione, o applicazione, usata per le diverse classi di funzioni, si concretizza in modi diversi e non è necessariamente legata all'indeterminatezza della funzione nelle sue singolarità. Esempi tipici di singolarità nei modelli matematici sono forniti dai momenti del tempo nei quali un processo si sviluppa a salti (discontinuità, ecc.); dai punti dello spazio ove si ha l'origine dei campi (gli elettroni nella teoria del campo elettromagnetico).

Un'applicazione a piú valori dell'insieme X nell'insieme Y è una applicazione $X \to \mathcal{P}(Y)$, dove $\mathcal{P}(Y)$ è l'insieme di tutti i sottoinsiemi di Y. Se $f(x) = Z \subseteq Y$, intuitivamente f associa ad $x \in X$ molti valori: tutti i possibili elementi $z \in Z$. Un tipico esempio di funzione a piú valori è la «funzione» $f(x) = \sqrt{x}$ dai numeri reali positivi a R: essa assume due valori in ogni punto $x \neq 0$. È, questo, un caso particolare di uno schema generale sul modo di formare funzioni analitiche a piú valori in una variabile complessa. Se una funzione è data in un intorno di un punto nel piano complesso mediante una serie di potenze, il processo del suo prolungamento analitico, che consiste nell'assegnarla mediante serie diverse in una catena di domini intersecantisi, può condurre a un valore diverso di questa funzione nel punto iniziale, quando la catena di domini si chiude su se stessa. In questo processo si possono anche individuare punti d'indeterminatezza della funzione. Pertanto, a partire da una funzione iniziale in un piccolo dominio, si giunge a una funzione da C in C (numeri complessi), la quale in generale è a piú valori e parziale. Cosí, per esempio, la funzione $\log x$ da C in C non è definita nel punto $x = 0$ (e solo in esso) e, per ogni $x \neq 0$, assume un insieme numerabile di valori, $\{\log_0 x + 2\pi\sqrt{-1}\, n\}$, dove n varia tra tutti i numeri interi, mentre $\log_0 x$ è uno di questi valori. Affinché questa funzione diventi un'applicazione ad un solo valore, bisogna cambiare il suo dominio di definizione. Questo importante processo è stato applicato la prima volta da B. Riemann [1851] e ha condotto alla fondamentale nozione di superfici di Riemann e, piú tardi, alla nozione generale di «varietà complesso-analitiche» (cfr. oltre). L'uso sistematico delle applicazioni parziali ha condotto negli ultimi decenni all'importante formalismo della teoria dei fasci (cfr. oltre).

In matematica hanno un maggiore significato non già le applicazioni arbitrarie d'insiemi, ma le applicazioni che concordano con queste o quelle strutture su essi. Però, prima di passare a descrivere degli esempi, è bene dire qualcosa sulla teoria delle potenze degli insiemi di G. Cantor [1879-83], dove la nozione di applicazione è apparsa per la prima volta nella sua forma «pura», consentendo in tal modo di porre tutto un complesso di difficili problemi matematici e filosofici.

Prima di tutto, la nozione di applicazione permette di formulare nel linguaggio della teoria degli insiemi la definizione esatta d'insieme finito. *Un insieme X si dice finito* (secondo Dedekind) *se ogni sua applicazione iniettiva in se stesso è contemporaneamente surgettiva.* Intuitivamente il senso di questa definizione è già stato chiarito: se un'applicazione f non incolla degli elementi, cioè, se la relazione di equivalenza E_f è l'identità, l'immagine $f(X)$ deve contenere tanti elementi quanti ne contiene X e perciò non può essere strettamente minore di X. Questo ragionamento perde significato per insiemi infiniti X (che cosa significa «tanti elementi quanti»?) e, infatti, l'applicazione $f : Z \to Z$, $f(x) = 2x$, dell'insieme dei numeri interi in se stesso è iniettiva, ma non surgettiva. Tale «paradosso dell'infinito» – «il tutto è uguale a una sua parte» – si discuteva da molto tempo e Dedekind propose [1888] di considerarlo semplicemente quale definizione dell'infinito nell'universo degli insiemi. (Negli ultimi tempi un paradosso analogo si è incontrato nella fisica delle particelle elementari. Un nucleone può generare una coppia virtuale nucleone-antinucleone e, se si tenta di ritenere i termini di questa coppia quali parti del nucleone, bisogna riconoscere che il nucleone contiene una sua parte propria isomorfa a se stesso. Per molti versi le particelle elementari vanno considerate quali «infiniti fisici»).

Fino a Cantor, però, non era noto che anche gl'insiemi infiniti si prestano a un confronto rispetto alla loro grandezza. A Cantor risale la serie seguente di definizioni. Due insiemi, X, Y, si dicono equipotenti se esiste un'applicazione biettiva $f : X \to Y$ (o, ciò che è lo stesso, una biezione $g : Y \to X$). In questo caso scriveremo card $X =$ card Y. Scriveremo card $X \leq$ card Y se X è equipotente a un sottoinsieme di Y. Diremo che X e Y sono confrontabili se card $X \leq$ card Y, oppure card $Y \leq$ card X. Infine, scriveremo card $X <$ card Y (la potenza di X è minore della potenza di Y) se card $X \leq$ card Y, ma non è vero che card $X \geq$ card Y.

Dagli assiomi di Zermelo-Fraenkel si può dedurre il seguente teorema di Cantor-Schröder-Bernstein-Zermelo:

a) *ogni due insiemi sono confrontabili; se contemporaneamente* card $X \leq$ card Y *e* card $Y \leq$ card X, *allora* card $X =$ card Y;
b) *sia* $\mathcal{P}(X)$ *l'insieme di tutti i sottoinsiemi di* X; *allora* card $\mathcal{P}(X) >$ card X;
c) *per ogni proprietà degli insiemi esiste un insieme di potenza minima con questa proprietà (se questi insiemi esistono in generale).*

Pertanto le potenze degli insiemi sono linearmente ordinate (cfr. oltre); la massima potenza non esiste; infine le potenze degli insiemi sono completamente ordinate. Queste proprietà della scala delle potenze furono scoperte da Cantor, benché alcune di esse siano state dimostrate piú tardi.

La scala delle potenze comincia con i numeri interi non negativi 0, 1, 2, 3, ..., che possono essere *definiti* quali potenze degli insiemi finiti [Whitehead e Russell 1962]. La prima potenza infinita è quella dell'insieme $N = \{0, 1, 2, 3, \ldots\}$ dei numeri interi non negativi, la cosiddetta potenza numerabile ω_0, oppure $\aleph_0$. La potenza infinita successiva strettamente maggiore, che si può costruire con il metodo di Cantor, è la potenza dell'insieme $\mathcal{P}(N)$. Essa si chiama po-

tenza del continuo e si indica con 2^{ω_0} in analogia con il caso degli insiemi finiti: se X contiene n elementi, $\mathcal{P}(X)$ contiene 2^n elementi.

Il fatto che la potenza di $\mathcal{P}(N)$ sia strettamente maggiore della potenza di N si dimostra con il famoso processo di diagonalizzazione di Cantor. Infatti, N è biettivo con la parte di $\mathcal{P}(N)$ costituita da tutti i sottoinsiemi con un solo elemento, $\{n\}$, dove $n \in N$. Se N fosse biettivo con tutto $\mathcal{P}(N)$, dovrebbe esistere una applicazione $f: N \to \mathcal{P}(N)$ con immagine $\mathcal{P}(N)$. Mostriamo che, per ogni f, l'immagine di f non può coincidere con $\mathcal{P}(N)$. Infatti, indichiamo con $M \subseteq N$ il sottoinsieme di N (eventualmente vuoto), costituito da tutti i numeri m tali che $m \notin f(m)$. Supponiamo che $M = f(m_0)$ per un opportuno $m_0 \in N$ e giungeremo a una contraddizione. A questo scopo consideriamo la posizione di m_0 rispetto ad M. Se $m_0 \in M$, in forza della definizione di M, $m_0 \notin f(m_0)$, cioè $m_0 \notin M$, poiché $f(m_0) = M$; questa è una contraddizione. Se, invece, $m_0 \notin M$, essendo $f(m_0) = M$, risulta $m_0 \notin f(m_0)$, ma allora $m_0 \in M$ per definizione di M; si ha di nuovo una contraddizione. Ciò dimostra il teorema di Cantor. È evidente che non si è mai usato il fatto che N sia l'insieme $\{0, 1, 2, \ldots\}$, cosí che la dimostrazione e il risultato sono applicabili ad ogni insieme al posto di N.

La questione relativa all'esistenza di potenze intermedie tra ω_0 e 2^{ω_0}, che si pone in modo naturale nell'ambito di queste idee, si chiama «problema del continuo». Dopo prolungati sforzi fu scoperto che l'ipotesi di Cantor, secondo la quale non vi sono potenze intermedie, è *indipendente* dagli altri assiomi della teoria degli insiemi. Il potente metodo per la dimostrazione dell'indipendenza, ideato da P. Cohen [1966], condusse a stabilire l'indipendenza di tutta una serie di altre affermazioni nella teoria degli insiemi, che si possono interpretare quali insolubilità di una serie di problemi relativi all'esistenza di applicazioni d'insiemi infiniti, con proprietà di questo o di quel tipo, nell'ambito dell'assiomatica assunta.

Tuttavia, la logica interna dello sviluppo della matematica e le esigenze delle scienze naturali, nella maggioranza dei casi, inducono a considerare non già qualsiasi applicazione, ma solo quelle che concordano con queste o quelle strutture (cfr. il paragrafo seguente).

3. *Applicazioni di strutture matematiche.*

In questo paragrafo si riporterà la descrizione di alcune fondamentali strutture matematiche e si chiarirà come si possa legare con ogni struttura su un dato insieme una classe di applicazioni di questo insieme che concordino con la struttura data.

3.1. Struttura di ordinamento.

Sia X un insieme. Un sottoinsieme $R \subseteq X \times X$ si chiama relazione binaria R su X. Al posto di $\langle x, y \rangle \in R$, per $x, y \in X$, spesso si scrive xRy. Una relazione binaria $\leq$ su X definisce una struttura d'ordinamento se da $x \leq y$ e $y \leq x$

segue che $x = y$ e se da $x \leq y$ e $y \leq z$ segue che $x \leq z$ (transitività). Il piú noto esempio di struttura di ordinamento è l'ordinaria relazione di disuguaglianza (non stretta) sui numeri reali, razionali, o interi; essa motiva la scelta della notazione generica $\leq$ per la relazione di ordinamento sugli insiemi astratti. Un altro esempio è la relazione $\subseteq$ sull'insieme $\mathcal{P}(X)$ dei sottoinsiemi di X. Un terzo esempio è la relazione card $x \leq$ card y su un certo insieme di potenze di insiemi. Nella nostra descrizione del modello di puro scambio in economia matematica (§ 1.6) ogni funzione di utilità $u_i : P \to R$ definisce una certa struttura di preordinamento $\leq_i$ sull'insieme P degli assortimenti di merci: $(x_1, \ldots, x_l) \leq (x'_1, \ldots, x'_l)$ se $u_i(x_1, \ldots, x_l) \leq u_i(x'_1, \ldots, x'_l)$, ma può non valere il primo assioma. Le strutture $\leq_i$, in generale, non coincidono. Non è difficile vedere che solo queste strutture sono essenziali nello studio dell'esistenza dello stato ottimale secondo Pareto; se si cambia u_i in modo da non cambiare $\leq_i$ (per esempio, se, invece di u_i, si prende $c_i u_i$ con $c_i > 0$), l'insieme degli stati ottimali non cambia. Vale la pena di notare che la validità reale di una congettura sulla transitività di una relazione di utilità $\leq_i$ è problematica: sono noti esperimenti nei quali un essere sottoposto a tre stimoli, X_1, X_2, X_3, preferisce X_1 allo stimolo X_2, X_2 allo stimolo X_3, ma preferisce X_3 allo stimolo X_1. Comunque sia, in una serie d'indagini sociali e psicologiche le strutture di ordinamento appaiono in modo naturale nelle seguenti circostanze. Posto che c'interessi l'intensità di questo o quel segnale di un fenomeno, allora accade spesso che si sia in grado di stabilire, per ogni coppia di fenomeni, a quale di essi corrisponda un segnale piú intenso, il che conduce a una struttura di ordinamento. La tendenza a descrivere in ogni caso questa intensità quantitativamente, cioè di stimarla con una funzione numerica, potrebbe portare ad introdurre un inutile elemento arbitrario non richiesto dall'essenza dei fatti.

Un insieme X con su esso una struttura di ordinamento si dice insieme ordinato. Un'applicazione $f : X \to Y$ di due insiemi ordinati si dice *monotona* se da $x_1 \leq x_2$ in X segue che $f(x_1) \leq f(x_2)$ in Y. Nell'esempio con le funzioni di utilità l'applicazione $u_i : P \to R$ è monotona rispetto al preordinamento $\leq_i$ su P e all'ordinaria disuguaglianza su R.

Qualche volta appare utile considerare applicazioni *strettamente monotone* d'insiemi ordinati. Sia X un insieme con una relazione di ordinamento $\leq$. Introduciamo una nuova relazione $<$ su X: scriveremo $x_1 < x_2$ se $x_1 \leq x_2$, ma $x_1 \neq x_2$. Un'applicazione $f : X \to Y$ si dice strettamente monotona se da $x_1 < x_2$ segue che $f(x_1) < f(x_2)$. La funzione $f(x) = kx$ da R in R è strettamente monotona per $k > 0$ e solamente monotona per $k = 0$. Ecco ancora un esempio: siano $Y \subseteq X$ un insieme e un suo sottoinsieme. L'applicazione $\mathcal{P}(X) \to \mathcal{P}(Y)$, che ad ogni sottoinsieme $Z \subseteq X$ associa la sua intersezione con Y, è monotona rispetto alla relazione $\subseteq$ su $\mathcal{P}(X)$ e $\mathcal{P}(Y)$. Se $Y \neq X$, essa non è strettamente monotona.

Le applicazioni monotone e strettamente monotone sono gli esempi piú semplici di applicazioni che concordano con le strutture. Passiamo ad altri esempi.

3.2. Strutture algebriche.

Le strutture algebriche sono quelle definite dalle diverse leggi di composizione. Sia X un insieme. *Si chiama legge di composizione interna binaria su X ogni applicazione* $f : X \times X \to X$. Esempi: $X = R$, oppure Z, $f(x, y) = x + y$, oppure xy; oppure $X = \mathcal{P}(Y)$, $f(x, y) = x \cup y$, oppure $x \cap y$. In questi esempi f soddisfa la seguente condizione di *associatività*: $f(f(x, y), z) = f(x, f(y, z))$ per tutti gli x, y, z. Un insieme X con una legge di composizione interna associativa si chiama monoide. Tale legge di composizione spesso si indica con l'ordinaria notazione di prodotto, scrivendo xy al posto di $f(x, y)$. Se X, Y sono due monoidi, un'applicazione $g : X \to Y$ si chiama omomorfismo di monoidi quando $g(x_1 x_2) = g(x_1) g(x_2)$ per tutti gli x_1, $x_2 \in X$. In altre parole, g deve concordare con le leggi di composizione su X, Y.

Sia X un monoide. Un elemento $e \in X$ si chiama unità (bilatera), se $ex = xe = x$ per ogni $x \in X$. Una applicazione $i : X \to X$ si chiama inversione (bilatera), se $i(x)x = xi(x) = e$ per ogni $x \in X$. Al posto di $i(x)$ spesso si scrive x^{-1}. Un monoide X con la sopraindicata unità $e \in X$ e con l'applicazione $i : x \to x^{-1}$ si chiama gruppo. Siano X, Y due gruppi. Un'applicazione $f : X \to Y$ si chiama omomorfismo di gruppi se essa è un omomorfismo di monoidi che applica l'unità di X nell'unità di Y e se $f(x^{-1}) = f(x)^{-1}$ per ogni $x \in X$. Un monoide, o un gruppo, X si dice commutativo se $xy = yx$ per tutti gli x, $y \in X$.

L'insieme R_+^* dei numeri reali positivi è un gruppo commutativo con il prodotto quale legge di composizione, con l'unità 1 e l'inversione $i(x) = x^{-1}$. L'insieme R di tutti i numeri reali è un gruppo commutativo con la somma quale legge di composizione, con l'unità 0 e l'inversione $i(x) = -x$. L'applicazione $\log : R_+^* \to R$ è un omomorfismo di questi gruppi. Essa, inoltre, è biettiva: tali applicazioni si dicono «isomorfismi».

Un esempio di gruppo non-commutativo è l'insieme $S(X)$ di tutte le applicazioni biettive di un certo insieme X in se stesso, cioè le permutazioni di X. Qui la composizione è la composizione delle applicazioni, l'applicazione identica $x \to x$ è l'unità. L'applicazione inversa f^{-1} è definita cosí: se $f(x) = y$, $f^{-1}(y) = x$. Il gruppo $S_n = S(\{1, ..., n\})$ si chiama gruppo simmetrico di n-esimo grado. Per $n \geq 3$ esso non è commutativo. Ad ogni elemento $s \in S_n$ può essere associato il numero $+1$, oppure -1, cosicché questa applicazione diventa l'omomorfismo $\varepsilon : S_n \to \{1, -1\}$ (il secondo è un gruppo con il prodotto quale legge di composizione). L'omomorfismo ε è definito univocamente dalla condizione $\varepsilon(s) = -1$ se s cambia posto esattamente a due numeri e lascia gli altri al loro posto.

I gruppi e i loro omomorfismi hanno un'importanza fondamentale nella matematica e nella fisica moderne. In particolare, essi forniscono i modelli fondamentali per la descrizione delle simmetrie degli oggetti, dei fenomeni e delle leggi della natura.

Gli esempi successivi di strutture algebriche sono i corpi e gli anelli. Un insieme K si chiama corpo se su esso sono date le seguenti strutture: α) la struttura di gruppo commutativo con legge di composizione $+$ e l'unità 0 su tutto K; β) la struttura di gruppo commutativo con unità 1 e legge di composizione «prodotto»: $(x, y) \to xy$ sull'insieme $K^* = K \setminus \{0\}$; e γ) queste strutture sono legate dalla legge distributiva $x(y+z) = xy + xz$ per tutti gli x, y, $z \in K$. Se in β) si richiede che K sia, rispetto al prodotto, solo un monoide, si ottiene la struttura di anello.

I numeri razionali Q, i numeri reali R e i numeri complessi C forniscono gli esempi piú noti di corpi. La nozione generale di corpo assiomatizza quelle proprietà di natura puramente algebrica. Esistono corpi con un numero finito di elementi; i cosiddetti «corpi di Galois». L'esempio piú semplice di tali corpi è $\{0, 1\}$, con le leggi di composizione $0+0 = 1+1 = 0$; $0+1 = 1$; $0 \cdot 0 = 0 \cdot 1 = 0$; $1 \cdot 1 = 1$. Questo corpo e corpi piú generali di Galois sono largamente usati nei moderni dispositivi per la trasmissione di dati e per l'elaborazione delle informazioni, dove sulla loro base sono progettati diversi sistemi di codificazione. Un'applicazione $f : K \to L$ di due corpi si chiama omomorfismo se essa è contemporaneamente un omomorfismo di K in L rispetto alla somma e di K^* in L^* rispetto al prodotto. Si può dimostrare che ogni omomorfismo di corpi è necessariamente iniettivo. Gli omomorfismi dei corpi sono studiati nella teoria di Galois e in geometria algebrica.

Siano K un corpo ed E un gruppo commutativo con la legge di composizione indicata con $+$ come in K e con l'unità 0 come nel gruppo additivo K. La struttura sul gruppo E di spazio vettoriale, o lineare sopra K è definita da una legge di composizione esterna $K \times E \to E$. L'immagine della coppia $\langle k, e \rangle$, dove $k \in K$, $e \in E$, in questa applicazione s'indica ke. Quest'applicazione deve soddisfare una serie di proprietà; in particolare, $(k_1 + k_2)e = k_1 e + k_2 e$, $k(e_1 + e_2) = ke_1 + ke_2$, $(k_1 k_2)e = k_1(k_2 e)$ per tutti i k, k_1, $k_2 \in K$; e, e_1, $e_2 \in E$.

L'esempio piú semplice di spazio lineare sopra K è il gruppo $E = K^n = K \times \times ... \times K$ (n volte), costituito dalle n-uple $(k_1, ..., k_n)$. La legge di composizione interna in E è: $(k_1, ..., k_n) + (k'_1, ..., k'_n) = (k_1 + k'_1, ..., k_n + k'_n)$. La legge di composizione esterna è: $k(k_1, ..., k_n) = (kk_1, ..., kk_n)$. Piú complessi e molto importanti nelle applicazioni sono gli spazi funzionali. Per esempio, lo spazio E delle funzioni da R in R è dotato della struttura di spazio vettoriale sopra R con le leggi $(f+g)(x) = f(x) + g(x)$ per tutte le f, $g \in E$ e per ogni $x \in R$ e $(cf)(x) = cf(x)$ per ogni $c \in R$, $f \in E$, $x \in R$.

Siano E_1, E_2 due spazi lineari sopra il corpo K. Un'applicazione $l : E_1 \to E_2$ si chiama applicazione lineare, o (omo)morfismo di questi spazi se essa è un omomorfismo di gruppi e, inoltre, se concorda con il prodotto esterno nel senso che è $l(ke) = k(le)$ per ogni $k \in K$, $e \in E$. Un'applicazione lineare di uno spazio E in se stesso si chiama operatore lineare su esso.

La nozione di applicazione lineare investe una parte notevole della matematica moderna e delle sue applicazioni. Essa è a fondamento di tutta l'analisi. La derivazione delle funzioni è un'operazione lineare; la sostituzione di piccole variazioni di una funzione con approssimazioni lineari è un principio generale nei modelli matematici, straordinario per la sua universalità e per la sua efficacia. L'integrazione delle funzioni è pure un'operazione lineare; la sua

base naturale è l'esistenza di un insieme di caratteristiche fisiche del tipo misura (area, volume, massa, energia), che sono additive quando due sistemi si riuniscono in uno e che, quando si cambia l'unità di misura, si comportano in modo naturale. Infine, la linearità dell'apparato matematico fondamentale della meccanica quantistica classica è basata sul principio della superposizione degli stati, la cui stessa formulazione sarebbe stata impossibile senza una preliminare individuazione matematica dell'arco di nozioni descritte. La stessa cosa si può dire della nozione dell'operatore lineare di una data grandezza fisica esaminata nel § 1.5.

Se gli stati di un dato sistema sono descritti da uno spazio lineare E, il gruppo G delle simmetrie di questo sistema è dotato di un omomorfismo di G nel gruppo degli operatori lineari biettivi su E. Questo omomorfismo si chiama rappresentazione lineare del gruppo G. Le rappresentazioni lineari formano una ricca classe di strutture ben studiate. Per esempio, l'atomo di un qualsiasi elemento è dotato di un gruppo naturale di simmetrie, cioè le rotazioni dello spazio attorno al nucleo dell'atomo. Risulta che le caratteristiche matematiche di diverse rappresentazioni di questo gruppo possono essere identificate con una serie di caratteristiche sperimentali degli elementi; in una certa approssimazione la tabella di Mendeleev può essere persino considerata quale tabella degli invarianti di diverse rappresentazioni del gruppo di rotazioni.

3.3. Spazio con misura e flussi.

Consideriamo il moto di un liquido in un certo volume; per esempio, il flusso dell'acqua nel letto di un fiume. Con buona approssimazione si può ritenere l'acqua incomprimibile. Seguiamo il moto di un certo volume d'acqua; per esempio, di una sfera di raggio 1 centimetro, individuata al momento iniziale nel flusso generale. Dopo un intervallo di tempo, diciamo di 1 secondo, essa si è mossa in basso lungo la corrente e si è deformata, ma il suo volume è rimasto lo stesso. Tali generi di fenomeni sono matematicamente modellati con le nozioni di «spazi con misura», di «applicazioni che conservano la misura» e di «flussi».

Si chiama spazio con misura un insieme X nel quale sia individuato un certo sottoinsieme $S \subseteq \mathcal{P}(X)$, cioè i sottoinsiemi misurabili di X, e sia data una funzione $\mu : S \to \{\text{numeri reali non negativi}\}$. Per ogni insieme misurabile $Y \in S$ il valore $\mu(Y)$ si chiama misura di Y. L'insieme S deve essere chiuso rispetto alle intersezioni e alle unioni numerabili dei suoi elementi, mentre la misura μ deve godere della proprietà additiva: la misura dell'unione di un insieme numerabile di elementi di S, a due a due disgiunti, è uguale alla somma delle misure di questi elementi. Nell'esempio del flusso del liquido, $\mu(Y)$ è semplicemente il volume di Y.

Un'applicazione $T : X \to X$ *conserva la misura* se, per ogni insieme misurabile $Y \subseteq X$, l'insieme $T(Y)$ è misurabile e $\mu(T(Y)) = \mu(Y)$. Si considererà qui solo il caso in cui T è biettiva e tutte le potenze di T conservano la misura. Un flusso su X è un insieme $\{T_t\}$ di applicazioni che conservano la misura,

indicizzate con i «momenti del tempo» $t \in R$ e tali che $T_{t_2} \circ T_{t_1} = T_{t_2+t_1}$. Nel caso del flusso di un liquido T_t applica la posizione di ogni particella del liquido al momento 0 nel luogo che essa occupa al momento t (l'insieme X è l'insieme dei punti del volume di liquido).

Le applicazioni che conservano la misura sono un altro esempio di applicazioni che concordano con le strutture.

Il ruolo di tali applicazioni nella fisica non è limitato all'esempio riportato all'inizio di questo punto. Il fatto è che, per una classe estesa di sistemi fisici che si evolvono nel tempo, nel loro spazio delle fasi esiste una misura invariante rispetto al flusso di questa evoluzione (teorema di Liouville per i sistemi hamiltoniani). Un esempio può essere il sistema solare, oppure, in una certa approssimazione, il sistema di molecole in un dato volume di gas. (Qui lo spazio delle fasi non coincide affatto con lo spazio occupato da questo volume; si tratta di uno spazio astratto, ogni punto del quale è costituito dall'insieme delle coordinate e delle velocità di tutte le molecole. La sua dimensione è molto grande; in condizioni normali 1 cm^3 di gas contiene circa 10^{20} molecole). Dalla conservazione della misura segue l'importante teorema di Poincaré: nell'evoluzione di un tale sistema, per quasi ogni stato iniziale, il sistema ritorna infinite volte a uno stato arbitrariamente vicino a quello iniziale. In particolare, se inizialmente tutto il gas era concentrato in una metà del volume, per esempio, diviso da una paratia, dopo che questa sia stata tolta, il gas, essendo lasciato a se stesso, si raccoglierà da solo infinite volte in quella metà del volume. La contraddizione di questa deduzione con la pratica dell'uomo si chiama, nei fondamenti della meccanica statistica, paradosso di Zermelo. Di solito esso viene risolto osservando che il tempo medio per tornare alla posizione iniziale in questo caso è cosí grande da superare tutto il tempo dell'esistenza dell'Universo. Ciò è legato all'enorme dimensione dello spazio delle fasi. Osservando il moto browniano di una singola piccola particella, ci si può convincere del suo ritorno periodico in un intorno della posizione iniziale.

Ancora un'altra importante applicazione nella fisica statistica delle applicazioni con misura invariante è legata ai cosiddetti teoremi ergodici. Essi motivano la coincidenza di medie «spaziali» e «temporali», cioè di diverse caratteristiche di un sistema, mediate rispetto al tempo della sua evoluzione, e di queste stesse caratteristiche, mediate rispetto a tutti i possibili stati del sistema in un fissato momento del tempo.

3.4. Continuità.

Una funzione $r : R(\text{tempo}) \to R(\text{distanza})$ che descriva il moto di un punto materiale gode della proprietà che le distanze in due punti vicini nel tempo sono pure vicine tra loro. In generale la maggioranza delle funzioni che modellano processi naturali è tale che piccole variazioni dei valori dell'argomento comportano piccole variazioni dei valori della funzione. Ciò conduce all'idea di continuità. (Costituiscono un'importante eccezione i domini singolari dello sviluppo di un processo, dove piccole variazioni dei parametri conducono a va-

riazioni molto rapide delle caratteristiche del processo: i modelli matematici di tali fenomeni forniscono la teoria delle singolarità di diversi generi, dei quali si tratterà in seguito in modo piú particolareggiato).

Una precisazione di queste idee conduce alla seguente definizione classica di funzione continua: *una funzione $f: R \to R$ si dice continua nel punto $x_0 \in R$ se, per ogni $\eta > 0$, esiste $\varepsilon > 0$ tale che, per $|x - x_0| < \varepsilon$, abbiamo $|f(x) - f(x_0)| < \eta$*: la variazione della funzione diventa piccola a piacere, se la variazione dell'argomento è sufficientemente piccola.

Questa definizione si estende evidentemente alle funzioni a valori reali in piú argomenti reali, alle funzioni in variabili complesse, alle funzioni-vettori, ecc.

Un'analisi ulteriore della nozione di continuità conduce a suddividerla in due nozioni: una si riferisce alla presenza di certe caratteristiche di «vicinanza» fra i punti del dominio X e del codominio Y dell'applicazione, cioè alle topologie su X e su Y; l'altra, alla conservazione di questa vicinanza in una applicazione continua. Nelle situazioni classiche il problema della topologia in R non nasce poiché esso entra a far parte delle proprietà usuali dei numeri reali.

La nozione generale di spazio topologico, cioè di un insieme con una topologia, è la seguente: *si chiama spazio topologico una coppia (X, T), dove $T \subseteq \mathcal{P}(X)$ è un certo insieme di sottoinsiemi di X, detti sottoinsiemi aperti di X, che soddisfa i seguenti assiomi:*

T_0) *$\emptyset \in T$ e $X \in T$;*

T_1) *per ogni punto $x \in X$ esiste un insieme $U \in T$ tale che $x \in U$.* [L'assioma T_1 è conseguenza di T_0, ma in questo contesto e per il seguito è utile indicare esplicitamente questa proprietà]*;*

T_2) *l'intersezione di ogni insieme finito di elementi di T è un elemento di T;*

T_3) *l'unione di ogni insieme di elementi di T è un elemento di T.*

Per esempio, nell'ordinaria topologia di R gli insiemi aperti sono esattamente quegli insiemi $U \subseteq R$ che, se contengono il punto $x \in R$, contengono l'intero intervallo $(x - \varepsilon, x + \varepsilon) \subseteq U$ se ε è sufficientemente piccolo. Analogamente, nell'ordinaria topologia dello spazio a n dimensioni R^n, un insieme $U \subseteq R^n$ è aperto se, e solo se, per ogni punto $\langle x_1, \ldots, x_n \rangle \in U$, esiste un $\varepsilon > 0$ tale che, per $|x_1 - y_1| < \varepsilon, \ldots, |x_n - y_n| < \varepsilon$, si ha $\langle y_1, \ldots, y_n \rangle \in U$.

In generale, siano (X, T_X) e (Y, T_Y) due spazi topologici. Un'applicazione $f: X \to Y$ si dice continua, o «morfismo di spazi topologici», se, per ogni insieme aperto $U_Y \in T_Y$, si ha $f^{-1}(U_Y) \in T_X$.

Dietro queste definizioni sono le seguenti considerazioni intuitive. La vicinanza dei numeri x, y di R è misurata dalla distanza tra loro, cioè dalla grandezza $|x - y|$; però questo non è il modo piú generale per introdurre la vicinanza, poiché è troppo strettamente legato alle idee usuali sulla retta reale. Invece, in generale basta limitarsi a dichiarare quali insiemi di punti debbano ritenersi «sempre piú vicini» a un punto x, e un'idea adeguata di ciò viene data dal «sistema di intorni aperti» di x, cioè dagli insiemi aperti $U \in T$ che contengono x. Un'applicazione $f: X \to Y$ è continua se punti vicini di X possono capitare solo in punti vicini di Y.

La maggior parte degli spazi topologici, che nell'analisi si presentano in modo naturale, soddisfano una condizione supplementare, il cui significato intuitivo consiste nel fatto che due punti distinti non possono essere «infinitamente vicini». Essa è formulata con espressioni diverse, piú o meno forti, tra le quali la piú usata è

T_4) (assioma di separabilità di Hausdorff): *per ogni due punti $x_1, x_2 \in X$, $x_1 \neq x_2$, esistono $U_1, U_2 \in T$ tali che $x_1 \in U_1$, $x_2 \in U_2$ e $U_1 \cap U_2 = \emptyset$.*

Le applicazioni continue di spazi separabili secondo Hausdorff sono dotate di molte proprietà che sono naturali per le funzioni classiche continue.

3.5. Varietà topologiche.

Fra i diversi spazi topologici sono particolarmente importanti gli spazi vettoriali R^n e C^n e gli spazi che *localmente* (in un intorno sufficientemente piccolo di ogni loro punto) sono strutturati in modo simile a R^n, o C^n.

Per formulare esattamente questa definizione introduciamo la nozione di «omeomorfismo». Si chiama omeomorfismo di due spazi topologici X, Y un'applicazione biettiva continua $f: X \to Y$ tale che l'immagine di ogni insieme aperto in X è aperta in Y e la controimmagine di ogni insieme aperto in Y è aperta in X (quest'ultima condizione segue dalla continuità di f). Per esempio, tutti gli intervalli aperti in R sono omeomorfi tra loro: l'omeomorfismo $(a, b) \to (0, 1)$ è dato dall'applicazione $f(x) = (x - a)/(b - a)$. Essi sono pure omeomorfi a tutta la retta R: l'omeomorfismo $(0, 1) \to R$ è dato dall'applicazione $f(x) = \log \log 1/x$. Analogamente si può dimostrare che tutte le sfere aperte in R^n sono omeomorfe tra loro e a tutto R^n.

Uno spazio topologico X si chiama varietà (ad n dimensioni) se, per ogni punto $x \in X$, esiste un suo intorno aperto U omeomorfo a R^n. Qui si sottintende che la topologia su U è definita dall'insieme di quei sottoinsiemi di U che sono aperti quali sottoinsiemi di X, cioè la topologia su U è *indotta* dalla topologia di X.

Nella fisica moderna sono di importanza fondamentale le varietà a quattro dimensioni, cioè i modelli matematici dello spazio-tempo. Il modello classico di Newton risponde allo spazio-tempo $R^4 =$ «spazio fisico» $\times$ «asse del tempo». Nel modello della teoria ristretta della relatività di Einstein lo spazio resta omeomorfo a R^4 (e conserva persino la struttura lineare), ma l'asse del tempo perde la sua posizione individuale. Infine, nei modelli della teoria generale della relatività lo spazio-tempo può appartenere ad una classe molto generale di varietà a quattro dimensioni che solo «in piccolo» conservano le proprietà di R^4.

3.6. Varietà differenziabili.

Lo spazio R^n è dotato in modo naturale di una struttura complementare che è data dall'insieme di tutte le funzioni $R^n \to R$ *differenziabili* rispetto a tutti gli argomenti. Una funzione $f(x_1, \ldots, x_n)$ si dice differenziabile in un intorno del

punto $(x_1^0, \ldots, x_n^0)$ se in questo intorno f può essere ben approssimata da una funzione lineare: $f(x_1, \ldots, x_n) = f(x_1^0, \ldots, x_n^0) + f_1' \cdot (x_1 - x_1^0) + \ldots + f_n' \cdot (x_n - x_n^0) +$ $+$ (termini infinitesimi d'ordine superiore rispetto a $\max(|x_1 - x_1^0|, \ldots, |x_n - x_n^0|)$); qui f_i' sono costanti dipendenti da $(x_1^0, \ldots, x_n^0)$. Se la funzione f è differenziabile in un intorno di ogni punto di R^n, essa si dice semplicemente differenziabile. In questo caso il coefficiente f_i', considerato quale funzione di $(x_1^0, \ldots, x_n^0)$, si indica

$$\frac{\partial f}{\partial x_i} : R^n \to R$$

e si chiama derivata parziale di f rispetto a x_i. Ripetendo questo processo si definiscono le derivate superiori

$$\frac{\partial^2 f}{\partial x_i \, \partial x_j} = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)$$

e cosí via. Se sono definite tutte le derivate

$$\frac{\partial^i f}{\partial x_1^{i_1} \ldots \partial x_n^{i_n}}$$

per $i_1 + \ldots + i_n = i \leq m$, la funzione f si dice m volte differenziabile. Una funzione differenziabile m volte per ogni m si dice infinite volte differenziabile, o liscia.

Sia X una varietà topologica ad n dimensioni. Un'applicazione $f : R^n \to X$ che definisca un omeomorfismo di R^n con un insieme aperto $U \subseteq X$ si chiama *carta* su X. In questo caso ogni punto $x \in U$ può essere dato dalle coordinate del punto $f^{-1}(x)$ in R^n; in questo senso assegnare le carte equivale ad assegnare su X un sistema locale di coordinate. Supponiamo che siano date due carte $f : R^n \to X$, $g : R^n \to X$, $f(R^n) = U$, $g(R^n) = V$. Allora nell'intersezione $U \cap V$ sono definiti due sistemi locali di coordinate; uno definito da f e l'altro definito da g. Esprimiamo le g-coordinate di un punto $x \in U \cap V$ mediante le sue f-coordinate: ciò fornisce un sistema di n funzioni in n variabili definito nel dominio $f^{-1}(U \cap V) \subseteq R^n$. Supponiamo che queste funzioni siano infinite volte differenziabili e che la stessa cosa sia vera per le espressioni delle f-coordinate mediante le g-coordinate. Allora le carte f e g si dicono compatibili. Sia ora (f_i) una certa famiglia di carte su X che ricoprano tutto X: $\bigcup_i f_i(R^n) = X$. Essa si chiama *atlante* su X se ogni due carte f_i, f_j sono compatibili.

Due atlanti (f_i) e (g_j) su X si dicono equivalenti se ogni carta f_i è compatibile con ogni carta g_j. Si può credere che la relazione di equivalenza fra gli atlanti si definisce effettivamente cosí.

La struttura di varietà liscia (o infinitamente differenziabile) su X si definisce con una certa classe di atlanti equivalenti su esso. Tutti questi atlanti si possono riunire in un atlante massimale.

Se nella definizione della compatibilità delle carte si sostituisce la differenziabilità infinita con la richiesta di essere differenziabili m volte, si definisce in modo analogo una varietà differenziabile m volte.

Siano X, Y due varietà lisce di dimensione m e n rispettivamente; $f : X \to Y$ sia un'applicazione continua. Allora in un intorno di ogni punto $x \in X$, usando le coordinate locali in x e in $f(x)$, possiamo considerare f come un'applicazione parziale di R^m in R^n. Perciò la teoria locale delle applicazioni di varietà differenziabili coincide con l'analisi classica reale. L'introduzione di una struttura globale nella forma di atlanti, che risale a H. Weyl che la ideò nel contesto delle superfici riemanniane, permette di usare i potenti mezzi della topologia nello studio di varietà globalmente non-omeomorfe a R^n e delle funzioni su esse. Per questo attualmente le varietà differenziabili sono l'oggetto fondamentale dell'analisi globale e della topologia differenziale. Esse appaiono quali spazi delle fasi dei sistemi fisici. In particolare, nella teoria generale della relatività i modelli spazio-tempo sono varietà differenziabili.

Due varietà lisce X, Y si dicono *diffeomorfe* se esiste un omeomorfismo $f : X \to Y$ che, in coordinate locali, sia dato mediante funzioni lisce dotate di funzioni inverse lisce. Una scoperta notevole degli ultimi decenni consiste nel fatto che esistono varietà molto classiche, per esempio le sfere, sulle quali possono essere introdotte strutture lisce sostanzialmente diverse tra loro. Per esempio, sulla sfera a sette dimensioni di tali strutture ve ne sono 28 [Milnor 1956].

3.7. Varietà analitiche.

Sul modello della definizione delle varietà differenziali si possono introdurre le varietà analitiche complesse (e reali). Sia $U \subseteq C^n$ un certo insieme aperto.

Una funzione $f : U \to C$ si dice analitica nel punto $\langle x_1^0, \ldots, x_n^0 \rangle \in U$ se essa è sviluppabile in una serie di potenze:

$$f(x_1, \ldots, x_n) = \sum a_{i_1 \ldots i_n} (x_1 - x_1^0)^{i_1} \ldots (x_n - x_n^0)^{i_n},$$

dove $a_{i_1 \ldots i_n} \in C$, $i_j \geq 0$, convergente in un intorno di questo punto. Una funzione f si dice analitica su U se ciò è vero per ogni punto di U.

Si chiama carta analitica su una varietà topologica X di dimensione $2n$ ogni omeomorfismo $f : U \to X$ della sfera aperta $U \subseteq C^n$ con un sottoinsieme aperto di X. Una carta analitica f fornisce su X le coordinate locali analitiche. Cosí come nel caso liscio, due carte si dicono compatibili se il passaggio dalle une alle altre coordinate è dato da funzioni analitiche. Dopo di che è chiaro come si definiscono gli atlanti analitici, la struttura di varietà analitica complessa su X e la nozione di isomorfismo analitico $f : X \to Y$.

Questa struttura è piú rigida della struttura di varietà differenziabile. Per esempio, la parte interna di un cerchio in C è diffeomorfa con tutto il piano complesso C, ma non è analiticamente isomorfa a C. Infatti, un'applicazione analitica di C in un cerchio è una funzione analitica limitata su tutto il piano, ma ognuna di tali funzioni è costante.

Le definizioni date di varietà lisce e analitiche mostrano la tecnica generale nell'usare le applicazioni (parziali) per incollare oggetti globali a modelli

locali di un dato tipo. Per concludere, è opportuno descrivere ancora una classe importante di strutture, dove questa stessa idea interviene in una variante un po' diversa.

3.8. Fasci.

Cominciamo con un caso particolare. Sia X una varietà analitica ed U sia un insieme aperto. Una funzione $f: U \to C$ si dice analitica se essa è analitica nelle coordinate locali in un intorno di ogni punto $x \in U$. Indichiamo con $\Gamma(U, O_X)$ l'insieme delle funzioni analitiche su U. La collezione di tutti i $\Gamma(U, O_X)$ gode delle seguenti proprietà:

1) se $U \subseteq V$, in modo naturale è definita l'applicazione-restrizione $r_{V,U}: \Gamma(V, O_X) \to \Gamma(U, O_X)$. Se $U \subseteq V \subseteq W$, si ha: $r_{W,U} = r_{V,U} \circ r_{W,V}$;
2) se $U = \cup U_i$ ed è data una funzione analitica $f \in \Gamma(U, O_X)$, la famiglia delle sue restrizioni $f_i = r_{U,U_i} f \in \Gamma(U_i, O_X)$ è definita univocamente da f e le restrizioni f_i, f_j concordano su $U_i \cap U_j$, cioè $r_{U_i, U_i \cap U_j}(f_i) = r_{U_j, U_i \cap U_j}(f_j)$;
3) al contrario, se $U = \cup U_i$ e se per ogni U_i è data una funzione analitica $f_i \in \Gamma(U_i, O_X)$, essendo per tutti gli i, j $r_{U_i, U_i \cap U_j}(f_i) = r_{U_j, U_i \cap U_j}(f_j)$, esiste allora $f \in \Gamma(U, O_X)$ con la proprietà $r_{U,U_i}(f) = f_i$ per ogni i.

L'assiomatizzazione di queste proprietà conduce alla definizione generale di fascio di insiemi $\mathfrak{F}$ su uno spazio topologico X. Dare un fascio $\mathfrak{F}$ significa dare gli insiemi $\Gamma(U, \mathfrak{F})$, uno per ogni insieme aperto $U \subseteq X$, e le applicazioni-restrizioni $r_{V,U}: \Gamma(V, \mathfrak{F}) \to \Gamma(U, \mathfrak{F})$ per ogni coppia d'insiemi aperti $U \subseteq V$. Questo insieme di dati $\{\Gamma(U, \mathfrak{F}), r_{V,U}\}$ deve soddisfare esattamente quelle proprietà che si sono formulate sopra per $\Gamma(U, O_X)$. Occorre solo sottolineare che, nella definizione generale di fascio, gli elementi $f \in \Gamma(U, \mathfrak{F})$ non sono necessariamente funzioni su U in qualunque senso della parola e le applicazioni-restrizioni $r_{V,U}$ devono essere assegnate a priori.

La struttura di fascio è uno dei mezzi di globalizzazione più generali attualmente noti. Essa, insieme con la nozione di topologia, è servita da punto di partenza per le successive generalizzazioni di A. Grothendieck.

Osserviamo ancora che gli insiemi $\Gamma(U, \mathfrak{F})$ possono essere dotati di strutture complementari, per esempio, della struttura di gruppo. Se tutte le applicazioni $r_{V,U}$ sono omomorfismi di gruppi, $\mathfrak{F}$ si chiama fascio di gruppi. Analogamente si possono definire i fasci di gruppi abeliani, i fasci di spazi lineari, ecc. Per esempio, il fascio delle funzioni analitiche su una varietà analitica è un fascio di anelli, poiché la restrizione delle funzioni è compatibile con la loro somma e il loro prodotto.

4. *Singolarità delle applicazioni.*

Nei §§ 1 e 2 si è già accennato al significato naturale-scientifico della nozione di singolarità di un'applicazione. Qui si riporteranno alcune situazioni

matematiche caratteristiche, nelle quali questa nozione si presta a precisazioni diverse.

In generale la nozione di *singolarità* di un'applicazione $f: X \to Y$ presuppone la descrizione preliminare di una certa proprietà locale dell'applicazione stessa, che, in ogni punto $x \in X$, f può, o non può godere, ma che f gode nella maggioranza dei punti di X (in uno, o in un altro senso della parola). I punti, dove questa proprietà non ha luogo, sono punti singolari per f. Al limite f può persino non essere definita nei suoi punti singolari; ciò significa, in verità, che f ha la forma $f: X_0 \to Y$, dove X_0 è il complemento in X dell'insieme dei punti singolari di f, mentre le singolarità si caratterizzano per questo o quel comportamento di f nei punti di approssimazione $x_0 \in X_0$ al punto singolare $x_1 \in X \setminus X_0$. Seguono alcuni esempi.

4.1. Discontinuità e spezzature.

Sia D un certo intervallo della retta reale R; $f: D \to R$ sia una funzione reale su D. Essa è continua nel punto $x_0 \in D$ se, per x tendente a x_0, il limite di $f(x)$ esiste ed è uguale a $f(x_0)$. Questa definizione è equivalente a quella data al § 3. Nel caso contrario f si dice discontinua nel punto x_0. Esempi semplici di discontinuità: $f(x) = 0$ per tutti gli $x \neq x_0$, $f(x_0) = 1$; $f(x) = 0$ per $x \leq x_0$, $f(x) = 1$ per $x > x_0$. Un esempio piú complesso: $f(x) = \operatorname{sen} 1/x$ per $x \neq 0$, $f(0) = a$, dove a è un numero arbitrario; qui $f(x)$ può non aver del tutto limite per x che si approssima a zero, oppure avere quale limite ogni numero compreso tra -1 e $+1$ (estremi compresi) a seconda di come x si approssima a zero.

Una funzione $f(x)$ può essere continua in tutto il suo dominio di definizione, ma differenziabile solo fuori di alcuni punti. La condizione di differenziabilità in un punto x_0 significa la presenza di una tangente definita al grafico di f nel punto $\langle x_0, f(x_0) \rangle$, oppure, piú formalmente, l'esistenza del limite di

$$\frac{f(x) - f(x_0)}{x - x_0}$$

per x che tende a x_0. La funzione $f(x) = |x|$ (cioè, uguale a x per $x \geq 0$ e a $-x$ per $x \leq 0$) è differenziabile in ogni punto $x \neq 0$; la sua derivata $\frac{df}{dx}(x) = f'(x)$ ($= 1$ per $x > 0$, -1 per $x < 0$) nel punto $x_0 = 0$ ha una discontinuità, mentre la stessa funzione $f(x)$, cioè il suo grafico, in questo punto è «spezzato».

Può accadere che attorno a un punto x_0 la stessa funzione e alcune sue prime derivate, $f'(x), \ldots, f^n(x)$, siano continue, mentre $f^{n+1}(x)$ sia discontinua in x_0. Ciò può avvenire quando si «congiungono» curve diverse; per esempio, prolungando un segmento di retta con un arco di circonferenza tangente alla retta nel punto di congiunzione: qui $n = 1$. Nel progettare una strada ferrata, per esempio, un tale punto può essere considerato singolare. Il grafico dello spazio percorso da un punto materiale ha una tale singolarità nel punto dove improvvisamente una forza comincia, o smette di agire. È, questo, il moto

di un razzo vicino al momento in cui si stacca il propulsore e il razzo comincia a muoversi lungo una traiettoria balistica.

4.2. Singolarità delle funzioni analitiche.

Sia D un certo insieme aperto del piano complesso C; per esempio, la parte interna di un cerchio. Ricordiamo che una funzione $f: D \to C$ si dice analitica in un punto $x_0 \in D$, se questo punto ha un intorno U tale che, per ogni $x \in U$, si ha

$$f(x) = \sum_{n=0}^{\infty} a_n (x - x_0)^n$$

($a_n \in C$; la serie converge). Se l'applicazione è analitica in tutti i punti $x \in D \setminus \{x_0\}$, dove $x_0 \in D$, mentre in un intorno opportuno del punto x_0 ovunque, ad eccezione che per $x = x_0$, essa è rappresentabile mediante una serie convergente della forma

$$f(x) = \sum_{n=-m}^{\infty} a_n (x - x_0)^n, \qquad m > 0, \ a_m \neq 0,$$

si dice che f ha un *polo* nel punto x_0. Un polo è una singolarità della funzione f. Il coefficiente a_{-1} si chiama residuo di f nel punto x_0. Se in condizioni analoghe f è rappresentata da una serie della forma

$$f(x) = \sum_{n=-\infty}^{\infty} a_n (x - x_0)^n$$

e $a_n \neq 0$ per infiniti $n < 0$, si dice che f ha una *singolarità* essenziale nel punto x_0.

Si hanno definizioni analoghe per le funzioni analitiche complesse in piú variabili.

Nella teoria delle funzioni complesso-analitiche si usa parlare anche di singolarità *tipo ramificazione*. La condizione $f(x) = \sqrt{x}$, dove per $\sqrt{x}$ si sceglie il valore positivo per i reali $x > 0$, insieme con l'esigenza della continuità, definisce in modo univoco una funzione analitica nel dominio infinito $D = C \setminus \{x | x < 0\}$. L'insieme $\{x | x < 0\}$ è un *taglio* di C lungo la parte negativa dell'asse reale. Quando x si approssima dal di sopra, o dal di sotto a uno stesso punto x_0 di questo taglio, $f(x)$ tende a limiti diversi («rami» di f), ma questi limiti sono tanto piú vicini l'uno all'altro (e a zero), quanto piú x_0 si avvicina a zero. Un'esatta definizione delle singolarità tipo ramificazione richiede che sia introdotta la nozione di «dominio di ricoprimento universale» per $D \setminus \{x_0\}$, sul quale f diventa un'effettiva applicazione (univoca).

Negli ultimi anni lo studio delle singolarità delle funzioni analitiche ha assunto una notevole importanza nella teoria dei processi di diffusione delle particelle elementari. Il modello matematico generale dei processi di diffusione parte dalle seguenti idee. In un modo o in un altro si genera un fascio di particelle con una energia sufficientemente elevata; per esempio, in un acceleratore.

Si fa interagire con le particelle di un bersaglio (il quale pure può essere un fascio) e si registrano le particelle che si formano quale risultato dell'interazione. Ogni atto dell'interazione viene considerato come una certa applicazione che associa allo stato del sistema prima dell'interazione (fascio, bersaglio) lo stato del sistema dopo l'interazione (prodotti dell'interazione). Sia l'uno che l'altro stato sono descritti da vettori in un certo spazio lineare, mentre le probabilità di passare dallo stato iniziale a quello finale sono descritte da una speciale funzione complesso-analitica, detta ampiezza di transizione, oppure diffusione. L'ampiezza di transizione è una funzione nelle energie e negli impulsi degli stati iniziali e terminali delle particelle del sistema. Le singolarità dell'ampiezza di transizione hanno un importante significato fisico. Per esempio, i poli dell'ampiezza di transizione corrispondono alle particelle stabili, o instabili presenti nel sistema. I residui in questi poli possono essere interpretati quali indici dell'intensità delle forze d'interazione connesse con queste particelle. L'ampiezza di transizione può avere anche singolarità, tipo punti di ramificazione. Le singolarità vicine a questi punti di ramificazione pure si prestano ad essere interpretate quali forze; piú precisamente, si tratta delle cosiddette «ampiezze assorbenti».

4.3. Singolarità delle applicazioni differenziabili.

Sia D un intorno dello zero, x_0, nello spazio R^n; $f: D \to R$ sia una funzione. Essa è m volte differenziabile nel punto $x_0 = (0, \ldots, 0)$ se, vicino ad esso,

$$f(x) = \sum_{i=0}^{m} f_i(x) + \varphi(x),$$

dove le $f_i(x)$ sono forme di grado i nelle coordinate locali $(x_1, \ldots, x_n)$, mentre $|\varphi(x)| \leq C \max_i |x_i|^{m+1}$. Il punto x_0 si dice regolare per l'applicazione f se il termine lineare $f_1(x)$ dell'applicazione è diverso da zero. Ciò significa che $f(x)$, vicino allo zero, è ben approssimata da un'applicazione lineare e può essere linearizzata mediante un'opportuna trasformazione (non-lineare) del sistema di coordinate. In caso contrario f ha una singolarità in questo punto, il cui tipo è definito dai termini successivi, $f_i (i \geq 2)$, dello sviluppo. Se $f_2(x) = \pm x_1^2 \pm \pm \ldots \pm x_n^2$, f ha la piú semplice singolarità «quadratica non degenere». I modelli matematici dei fenomeni dove tali singolarità risultano importanti sono assai complessi; essi sono oggetto della teoria delle catastrofi di R. Thom [1972]. Con l'aiuto di tali modelli si tenta di descrivere le biforcazioni dei fenomeni a scatti, quando una piccola differenza nelle condizioni iniziali di sviluppo di un processo può condurre a un corso bruscamente diverso del processo stesso, simile, per esempio, alla diversificazione dei tessuti contigui di un embrione in sviluppo, oppure alla polarizzazione di una società omogenea sotto l'influenza di fattori politici, ecc.

Sempre negli ultimi tempi, idee analoghe sono discusse in cosmologia in ordine alla violazione di diverse simmetrie delle leggi fisiche. Si presume che,

benché le stesse equazioni fondamentali della fisica siano dotate di simmetrie di genere diverso, gli scostamenti dalle loro soluzioni osservate, che si realizzano effettivamente, siano dovuti al fatto che, all'inizio dello sviluppo dell'universo, la soluzione simmetrica sia risultata instabile e che, sotto l'influenza di fluttuazioni casuali, stia «cadendo» in una delle direzioni di sviluppo stabili, ma asimmetriche.

5. *Applicazioni dal punto di vista dell'intuizionismo e del costruttivismo.*

Come abbiamo piú volte sottolineato, la nozione di applicazione generalmente accettata nella matematica moderna è strettamente legata alla nozione di insieme e alle convenzioni fondamentali relative al linguaggio della teoria degli insiemi. Questa situazione si è venuta affermando storicamente in modo tutt'altro che subitaneo. Dopo Newton e Leibniz, fondatori dei principî dell'analisi matematica moderna, per molto tempo in matematica si sono considerate solo quelle applicazioni assegnate descrivendo in modo esplicito il modo di calcolarne i valori (con precisione arbitraria), o, al limite, quelle descritte con una *formula* esplicita in questo o quel frammento simbolico del gergo matematico. Per esempio, una funzione $R \to R$ poteva essere data dalle formule $y = x^2$, $y = \operatorname{sen} x$, oppure

$$y = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

L'idea di applicazioni «arbitrarie» si è formata parallelamente a quella di insiemi «arbitrari». Nel corso di questo sviluppo si superò il secolare preconcetto contro l'uso in matematica della nozione d'infinito attuale. Soltanto dopo che furono posti i fondamenti del linguaggio degli insiemi divenne possibile una precisa definizione della nozione generale di applicazione.

Da questo punto di vista le applicazioni, descritte da queste o quelle classi di formule, o di testi, in questo o quel linguaggio, possono essere considerate come uno speciale oggetto di studio. Qualche volta tale studio può confinare con la problematica della logica matematica, poiché presuppone un'esatta descrizione preliminare del linguaggio nel quale vengono descritte le funzioni di una data classe. Cosí è per la nozione di funzione algoritmicamente calcolabile. In altri casi questo linguaggio veniva individuato in modo sufficientemente chiaro quale frammento del gergo matematico generale. Cosí si può descrivere esplicitamente la classe delle funzioni (localmente) analitiche, definite mediante serie di potenze convergenti, oppure la classe delle soluzioni delle equazioni differenziali a coefficienti costanti, ecc.

Esistono scuole di specialisti dei fondamenti della matematica (l'intuizionismo, il costruttivismo e loro sfumature), le quali ritengono illegittime le nozioni generali d'insieme infinito e di applicazione arbitraria assunte nel metodo assiomatico. A ciò è legata la critica a molti modi di assegnare una funzione usati nella matematica teoretico-insiemistica; per esempio, all'assioma della scelta. In particolare, in diverse varianti della matematica costruttivistica

si considerano solo quegli oggetti che possono essere esplicitamente costruiti, in un senso ben definito della parola, mediante mezzi finiti precedentemente descritti. Queste «matematiche costruttive», di regola, possono essere, a loro volta, interpretate in termini teoretico-insiemistici, dopo di che essere considerate quali sottoteorie di classi speciali d'insiemi e di applicazioni vicine, nello spirito, alla teoria della calcolabilità algoritmica. Tuttavia gli intuizionisti e i costruttivisti filosoficamente piú coerenti (Brower, Heyting, A. A. Markov) non accettano tale interpretazione.

Le concezioni intuizionistiche della matematica possono essere suddivise, grosso modo, in due componenti: una negativa e l'altra positiva. La parte negativa è un sistema di argomentazioni a favore della rinunzia alla nozione di infinito attuale; contro l'uso dei principî della logica ordinaria applicati a questo infinito e, come corollario, a favore della rinunzia alla maggior parte delle definizioni, delle costruzioni e dei risultati che si hanno in matematica, tutto ciò essendo ingiustificato e incomprensibile. La parte positiva è un sistema di principî e di risultati proposti in sostituzione della matematica classica rifiutata. Questa parte si presta difficilmente ad una esposizione in un linguaggio vicino a quello classico, poiché presuppone la rinunzia a quelle stesse abitudini fondamentali di un matematico con una cultura ordinaria. Cosí, per un intuizionista un'argomentazione per una dimostrazione è quella che in ogni suo passaggio è «evidente» e non quella che in ogni suo passaggio risale a un sistema di assiomi e di regole d'inferenza precedentemente convenuti. L'«evidenza» è una nozione soggettiva, il cui significato varia da persona a persona e per tempi diversi Cionondimeno, gli intuizionisti assumono quali nozioni matematiche primitive quelle che «sono chiare per ogni essere umano normale e persino per i piccoli fanciulli» [Heyting 1956] e le nozioni di «essenza» e «possibilità di ripetere infinitamente questa essenza» quali sorgenti della nozione di numero naturale. Le proprietà elementari dei numeri naturali si manifestano mediante una «semplice considerazione». Non appena questi concetti siano assunti, l'aritmetica dei numeri interi e razionali nella variante degli intuizionisti risulta parallela a quella classica, ad eccezione del linguaggio (che per gli intuizionisti non è soggetto a precisazioni troppo rigide) e della semantica che, in generale, è difficile confrontare. Forti divergenze con la matematica classica cominciano a livello dei numeri reali e delle funzioni su loro. Nella matematica classica un numero reale può essere dato, per esempio, quale classe di successioni equivalenti di Cauchy di numeri razionali. Ma già nella nozione di «successione di Cauchy» è inclusa quella di applicazione dei numeri interi nei razionali. Gli intuizionisti sostituiscono questa nozione, che si appella all'insieme attualmente infinito dei numeri interi, con la nozione di «successione che si estende infinitamente». «Con ciò è indifferente in qual modo siano definiti i termini della successione; se con una legge, o con una libera scelta, o a sorte, o in qualsiasi altro modo» [Heyting 1956]. Pertanto il primo esempio di applicazione «arbitraria» fra insiemi infiniti riscontrabile nell'intuizionismo è legato all'idea del tempo fisico, che è incompleto e che non si presta ad essere considerato quale unico insieme di tutti i suoi momenti.

Questa idea si estende gradatamente sempre piú e si allarga il dominio della sua applicabilità. Cosí, per un intuizionista, l'affermazione, secondo la quale un numero reale ρ, precedentemente definito, non è razionale, può essere la constatazione del fatto che, al momento di questa enunciazione, la razionalità di ρ non è stata ancora dimostrata. Ciò diverge fortemente con il linguaggio della matematica classica e costringe all'uso di diverse precauzioni linguistiche affinché si possa distinguere la precedente enunciazione da quella con il seguente significato: «supponendo la razionalità di ρ, giungo in un modo o in un altro a una contraddizione». Per il matematico classico qui vi è una mescolanza di due diversi livelli di descrizione: un livello di «matematica vera e propria» e un livello di «descrizione di questa matematica», ma un intuizionista non è d'accordo con questo punto di vista.

Non è qui il caso di entrare in altri particolari. Da quanto si è detto è chiaro che le idee sugli oggetti matematici fondamentali, in particolare sulle applicazioni, che si hanno nella matematica classica e nell'intuizionismo sono sostanzialmente inconfrontabili. Ciò nonostante è possibile ed utile un'analisi sufficientemente formalizzata, o produrre modelli delle concezioni intuizionistiche nell'ambito della matematica classica. Benché ciò, secondo gli intuizionisti, non possa riflettere adeguatamente la sostanza delle loro idee, tali modelli risultano fruttuosi nella matematica classica e sono legati all'analisi delle strutture che in essa sorgono in modo indipendente. Cosí, nella costruzione dei modelli non-standard della teoria degli insiemi, con l'aiuto dei quali si dimostra l'insolubilità di una serie di problemi classici, si può scorgere la classica realizzazione delle idee intuizionistiche sulla «formazione indeterminata» delle successioni e sulla nozione relativa di «verità» che sono simili alla «verificabilità a un dato momento». A una tale comprensione si presta il metodo del forcing secondo Cohen nella teoria assiomatica degli insiemi.

Le concezioni intuizionistiche, a loro volta, furono sottoposte a critica da parte tanto dei seguaci della matematica classica, quanto dei suoi avversari che richiedono, però, una precisazione delle nozioni fondamentali dell'intuizionismo. A questi ultimi appartengono le diverse scuole di costruttivisti. Per i costruttivisti della scuola di A. A. Markov la nozione fondamentale della matematica è quella di «oggetto costruttivo» che, il piú delle volte, appare quale nozione di «testo in un alfabeto finito». La nozione di applicazione interviene quale nozione di «algoritmo» nel significato precedentemente chiarito. Un algoritmo elabora testi in altri testi. Un numero naturale n, per esempio, appare quale testo di n segni $\|...|$, disposti in fila e compresi tra segni speciali di «spaziatura» all'inizio e alla fine. Un'applicazione dell'insieme dei numeri naturali in se stesso è un algoritmo che elabora tali testi in testi dello stesso tipo. Un algoritmo, a sua volta, può essere considerato quale oggetto costruttivo, cioè un testo, scritto magari in un alfabeto piú grande. Pertanto sono possibili algoritmi applicabili ad altri algoritmi, cioè dei «funzionali», ecc. Sviluppando conseguentemente queste idee si può giungere a varianti costruttive di grandi frammenti della matematica.

Esaminando questo punto di vista bisogna tener presente la seguente circostanza. I risultati finiti delle teorie matematiche appaiono spesso effettivamente quali descrizioni di procedure di calcolo di funzioni di questo, o di quel tipo, d'invarianti numerici, ecc. Possono essere esempi gli algoritmi, o i programmi dei calcolatori che permettono di risolvere le equazioni della fisica matematica, di produrre le tavole delle funzioni speciali, ecc. In questo senso i risultati matematici «tangibili» sono affermazioni su oggetti costruttivi verificabili costruttivamente. Tuttavia, può accadere che, con metodi puramente costruttivi, non si possano riconoscere proprietà importanti di oggetti costruttivi; per esempio, la coincidenza di due serie di oggetti costruttivi assegnate con costruzioni diverse. Se f e g sono due funzioni calcolabili con argomenti e a valori naturali dati, diciamo, con due programmi per un calcolatore, può capitare che sia $f(n)=g(n)$ per ogni n e che, tuttavia, ciò non possa essere dimostrato senza ricorrere alla nozione d'infinito attuale (vale a dire, ai metodi della teoria analitica dei numeri). Ciò è legato al teorema di Gödel sulla incompletezza; un'opportuna estensione del linguaggio e dei suoi mezzi deduttivi può rivelare delle affermazioni, che si formulano nel vecchio linguaggio e sono vere, ma la cui veridicità non si presta ad essere stabilita con mezzi deduttivi del vecchio linguaggio.

Tale genere di considerazioni, oltre ad argomenti puramente filosofici, conduce la maggioranza dei matematici a rifiutare le limitazioni costruttivistiche quali basi accettabili per i fondamenti della matematica. Naturalmente ciò non diminuisce l'importanza dello studio sistematico degli insiemi e delle applicazioni costruttive nell'uno o nell'altro senso della parola.

Per dare un'idea della natura dei limiti dei mezzi costruttivi si farà l'esempio di un problema che è, per principio, irrisolubile con questi mezzi. Ci si baserà su alcune nozioni della teoria generale di programmazione per i calcolatori a cifre.

A un centro di calcolo pervengono programmi in un linguaggio algoritmico (tipo Algol). Essi vengono introdotti nel calcolatore, dove un programma speciale, il codificatore, li traduce in programmi nel linguaggio dei codici di macchina; solo dopo, un tale programma è accettato dal calcolatore e serve per un calcolo reale. Alcuni programmi nel linguaggio algoritmico contengono errori, diciamo sintattici; il codificatore reagisce ad alcuni errori fermandosi e segnalandoli al programmatore.

Sembra naturale porre il problema di scrivere un codificatore che in tutti i casi distingua i programmi regolari da quelli irregolari.

Precisiamo il problema. Fissiamo un linguaggio algoritmico L; il suo alfabeto è costituito da un numero finito di segni. Ogni successione finita di segni di questo alfabeto si chiama testo. Un testo si dice programma regolare se la sua traduzione nel linguaggio dei codici di macchina è possibile e se, dopo aver introdotto questo programma, la sua traduzione fa funzionare il calcolatore per un tempo potenzialmente infinito e conseguentemente produce in uscita una tabella della forma $(1, f(1))$; $(2, f(2))$; $(3, f(3))$; ..., dove gli $f(n)$ sono numeri naturali. In altre parole, i programmi regolari calcolano funzioni con argomenti naturali e a valori naturali.

Chiamiamo CODIFICATORE (C) un programma nel linguaggio L tale che, per ogni testo in ingresso, dopo aver funzionato, dia in uscita 1, se il testo è un programma regolare, e 0 nel caso contrario. I programmi regolari, cosí come tutti i testi, sono oggetti costruttivi, e il C è un mezzo costruttivo per la loro elaborazione.

Dimostriamo che il C non esiste. Supponiamo il contrario e, usando il C, scriviamo un certo programma regolare che chiameremo PROGRAMMA PRINCIPALE (PP). In un modo qualsiasi ordiniamo i simboli del linguaggio L; chiamiamolo ordinamento alfabetico. Corrispondentemente ordiniamo i testi nel linguaggio L in ordine alfabetico. Dopo di che supponiamo che il PP funzioni cosí:

1) uno speciale sottoprogramma del PP costruisce uno dopo l'altro tutti i testi nel linguaggio L in ordine alfabetico;
2) uno dopo l'altro i testi costruiti vengono assegnati in ingresso al C la cui uscita, 0 o 1, viene elaborata dal PP nel modo seguente. Se è 0, il testo non è un programma regolare e, in tal caso, il PP assegna in ingresso al C il testo successivamente costruito. Se è 1, il testo è un programma regolare per calcolare la funzione f_n, dove n è l'indice ordinale nella successione dei programmi regolari, e, in tal caso, il C traduce questo testo in codici di macchina; il programma cosí ottenuto viene fatto funzionare esattamente fino a calcolare il valore $f_n(n)$; a questo valore viene aggiunto 1 e all'uscita del PP si ha la coppia di numeri $(n, f_n(n)+1)$.

È chiaro che il PP è un programma regolare. Perciò la funzione $F(n)=f_n(n)+1$ dev'essere contenuta nella successione $(f_1, f_2, f_3, \ldots)$ delle funzioni calcolabili mediante programmi regolari, le quali tutte sono generate, l'una dopo l'altra, dal PP. Ma ciò porta a una contraddizione, poiché F evidentemente è diversa da ogni f_i quando in entrambe si ha quale argomento i. Pertanto il C non esiste.

Si noti che abbiamo richiesto che il C dev'essere scritto nello stesso linguaggio L dei programmi regolari; altrimenti il PP non potrebbe far parte dei programmi regolari in questo linguaggio e il ragionamento non avrebbe retto. Se il linguaggio L è sufficientemente povero, mentre il linguaggio del codificatore è sufficientemente ricco, il riconoscimento dei programmi regolari qualche volta risulta possibile. Ma ciò significa semplicemente che noi consapevolmente restringiamo la classe delle funzioni calcolabili considerate proprio per raggiungere questo scopo; se il linguaggio, nel quale è scritto il codificatore, è già dato, diventa naturale scrivere anche i programmi per il calcolo in questo linguaggio, usandolo in tutta la sua completezza.

Per concludere, va osservato che le discussioni sulla sostanza e sulla legittimità delle diverse idee matematiche sono essenzialmente filosofiche e non possono essere risolte in modo univoco. La questione dell'accettabilità di questa o quella considerazione viene risolta da ogni matematico a partire da profonde convinzioni psicologiche, sulle quali possono influire l'educazione ricevuta, il temperamento e persino il chimismo individuale dei processi biologici. Con

certezza si possono constatare solo due circostanze. Prima di tutto, la critica intuizionistica e costruttivistica alla matematica classica e moderna finora non pare convincente per la maggioranza degli specialisti di matematica, pura o applicata che sia. Inoltre, tutte le concezioni positive che vengono proposte in sostituzione di quelle classiche restringono fortemente l'arco di nozioni che si propone di ritenere legittime, e l'arco di affermazioni che si ritiene abbiano significato e siano vere. Ciò va contro i bisogni che s'incontrano nella maggioranza delle applicazioni; in particolare, nelle scienze fisiche, dove si richiedono mezzi sempre piú nuovi per riconoscere la verità e si usano formalismi matematici sempre piú nuovi che non stanno nel letto di Procuste di concezioni precedentemente preparate.

6. *Categorie e funtori.*

Il ruolo eccezionale delle applicazioni, quale nozione matematica fondamentale quanto la nozione di insieme e la nozione di struttura da esse derivate, è riflesso nella teoria delle categorie, formatasi negli ultimi due decenni. Inizialmente, le sue nozioni fondamentali si generarono quali assiomatizzazioni di una serie di situazioni simili in topologia e nell'algebra, particolarmente quella omologica. Con lo sviluppo del suo apparato e delle sue applicazioni alla geometria algebrica, all'analisi funzionale, alla logica, alla teoria dei numeri, ecc., il linguaggio delle categorie ha assunto gradatamente lo status di linguaggio fondamentale e persino alternativo rispetto al linguaggio della teoria degli insiemi.

Al livello attuale è meglio immaginare la situazione nel senso che la descrizione categoriale degli oggetti matematici è duale rispetto alla loro descrizione strutturale. Dal punto di vista della teoria delle strutture un gruppo, o uno spazio topologico è *un insieme dotato di una determinata struttura*, che è, rispettivamente, una legge di composizione che soddisfa gli assiomi gruppali, o un insieme dei suoi sottoinsiemi aperti che soddisfano gli assiomi della topologia. Dal punto di vista della teoria delle categorie un gruppo, o uno spazio topologico X è *un oggetto della categoria di tutti i gruppi* (rispettivamente di tutti gli spazi topologici), considerato insieme a tutti i suoi omomorfismi in tutti i gruppi e a tutti gli omomorfismi di tutti i gruppi in esso medesimo (rispettivamente, insieme a tutte le applicazioni continue che hanno X quale dominio, o codominio). Si può dire che la teoria delle categorie realizza un approccio «sociologico» all'oggetto matematico, nel quale quest'ultimo viene considerato non come individuo, ma come membro di una comunità di suoi simili.

In un certo senso, il punto di vista categoriale su una data struttura è tanto piú efficace quanto piú numerosi sono gli oggetti con tale struttura; infatti, la teoria di tali oggetti individuali, quale l'anello dei numeri interi, contiene molti risultati che forse non sono descrivibili nell'ambito di un approccio categoriale. La stessa osservazione vale nella maggior parte dell'analisi classica.

Si esporranno ora le nozioni fondamentali del linguaggio delle categorie.

Una categoria C è costituita dai dati seguenti:

a) *Un insieme, o una classe ObC, i cui elementi si chiamano oggetti della categoria C.*

b) *Per ogni coppia ordinata $\langle X, Y\rangle$ di oggetti di C è dato un insieme (eventualmente vuoto)* $\mathrm{Hom}(X, Y)$ *(oppure* $\mathrm{Hom}_C(X, Y)$*), i cui elementi si chiamano morfismi dall'oggetto X nell'oggetto Y.*

Al posto di $\varphi \in \mathrm{Hom}(X, Y)$ spesso si scrive $\varphi : X \to Y$, oppure $X \xrightarrow{\varphi} Y$. Qualche volta i morfismi si chiamano frecce; X è l'origine della freccia $\varphi : X \to Y$, mentre Y è la sua fine. Nella categoria ogni freccia ha una origine e una fine definite univocamente. L'insieme, o la classe $\bigcup \mathrm{Hom}(X, Y)$ (dove l'unione è estesa a tutte le coppie X, Y della categoria C) qualche volta si indica con $\mathrm{Mor}\, C$.

c) *Per ogni terna ordinata $\langle X, Y, Z\rangle$ di oggetti di C è data una applicazione*

$$\mathrm{Hom}(X, Y) \times \mathrm{Hom}(Y, Z) \to \mathrm{Hom}(X, Z).$$

Alla coppia di morfismi $\varphi : X \to Y$, $\psi : X \to Y$ essa associa il morfismo $X \to Z$, indicato con $\psi \circ \varphi$, oppure con $\psi\varphi$ e detto composizione di φ e ψ.

Questi dati devono soddisfare i due assiomi seguenti:

A) *Per ogni oggetto $X \in ObC$ esiste un (unico) morfismo identità* $\mathrm{id}_X : X \to X$, *per il quale* $\mathrm{id}_X \circ \varphi = \varphi$ *e* $\varphi \circ \mathrm{id}_X = \varphi$ *ogni volta che le composizioni di* id_X *con il morfismo φ sono definite, cioè quando φ ha la forma $Y \to X$, oppure, rispettivamente, $X \to Y$.*

B) *Qualunque siano $\varphi : X \to Y$, $\psi : Y \to Z$ e $\chi : Z \to U$, abbiamo:* $(\chi\psi)\varphi = \chi(\psi\varphi)$.

Il morfismo $\varphi : X \to Y$ si dice inverso sinistro del morfismo $\psi : Y \to X$ se $\varphi \circ \psi = \mathrm{id}_Y$; inverso destro se $\psi \circ \varphi = \mathrm{id}_X$. Un morfismo $\varphi : X \to Y$ si dice isomorfismo se è dotato sia dell'inverso sinistro che dell'inverso destro.

È da notare che una formalizzazione della teoria delle categorie comunemente accettata, indipendente dalla teoria formale degli insiemi, per ora non esiste. La nozione di insieme o di classe figura già nella definizione degli insiemi degli oggetti e dei morfismi delle categorie. Non si può escludere che ciò sia solo un dato della tradizione (la nozione di categoria si è formata nel periodo in cui hanno dominato e continuano a dominare le concezioni teoretico-insiemistiche). Tuttavia l'autore è incline a credere che, essendo la teoria delle categorie cosí penetrata dalla matematica teoretico-insiemistica, è improbabile ch'essa possa diventare la base per una costruzione radicalmente nuova dei fondamenti della matematica. La questione del rapporto fra le concezioni teoretico-categoriali e quelle teoretico-insiemistiche è già stata affrontata in precedenza; si pensa ch'esse forniscano dei modi di descrivere gli oggetti matematici tra loro complementari e che la rinunzia di ognuno di questi modi faccia diminuire la completezza dell'informazione ottenuta.

Si descriveranno ora alcuni esempi di categorie. Essi si dividono naturalmente in tre gruppi che risalgono alle consuetudini di tempi diversi.

PRIMO GRUPPO DI ESEMPI. In questo gruppo di esempi gli oggetti sono insiemi, dotati di una struttura di questo o quel tipo, mentre i morfismi sono applicazioni di questi insiemi che concordano, in un determinato senso della parola, con le strutture (cfr. § 3).

Categoria degli insiemi. Tutti gli insiemi dell'universo di Zermelo-Fraenkel formano gli oggetti di una categoria, i cui morfismi sono le applicazioni qualsiasi degli insiemi. Si possono considerare diverse sottocategorie di questa categoria: per esempio, la categoria degli insiemi finiti, oppure degli insiemi non piú che numerabili.

In generale, una categoria C' si dice sottocategoria della categoria C, se $ObC' \subseteq ObC$ e se, qualunque siano X, $Y \in ObC'$, si ha $\mathrm{Hom}_{C'}(X', Y') \subseteq \mathrm{Hom}_C(X, Y)$. Questa sottocategoria si dice completa se $\mathrm{Hom}_{C'}(X', Y') = \mathrm{Hom}_C(X, Y)$ per tutti gli X', $Y' \in C'$.

Categoria degli spazi topologici. Sono oggetti di questa categoria tutti i possibili spazi topologici e i morfismi sono tutte le loro applicazioni continue.

Categoria dei gruppi. Gli oggetti di questa categoria sono tutti i possibili gruppi e i morfismi sono gli omomorfismi di gruppi. Ricordiamo che, se G (rispett. H) è un gruppo e la composizione degli elementi $x, y \in G$ (rispett. $\in H$) s'indica xy, allora l'applicazione $f : G \to H$ si chiama omomorfismo, se $f(xy) = f(x)f(y)$ per tutti gli $x, y \in G$.

Categoria dei gruppi abeliani. Gli oggetti di questa categoria sono tutti i gruppi G con legge di composizione commutativa ($xy = yx$ per tutti gli $x, y \in G$), mentre i morfismi sono gli omomorfismi di gruppi. La categoria dei gruppi abeliani è una sottocategoria completa della categoria di tutti gruppi.

Categoria dei gruppi topologici. Un gruppo topologico è un insieme G dotato di due strutture; una di spazio topologico e l'altra di gruppo, dove la composizione $\langle x, y\rangle \to xy$ e l'applicazione elemento inverso, $x \to x^{-1}$, sono continue. Un morfismo di gruppi topologici $G \to H$ è un'applicazione continua che è un omomorfismo di gruppi. Per esempio, i numeri reali positivi R_+^* e tutti i numeri reali R sono gruppi topologici rispetto, rispettivamente, al prodotto e alla somma, mentre l'applicazione $\log : R_+^* \to R$ definisce un loro morfismo.

Categoria degli insiemi ordinati. Ricordiamo che una struttura di ordinamento su un insieme X è data da una relazione binaria $x \leq y$ fra gli elementi di X. Scriviamo $x < y$, se $x \leq y$, ma non è vero che $x = y$. Per esempio, l'ordinaria relazione di disuguaglianza definisce una struttura di ordinamento negli insiemi R e Z. Una applicazione $f : X \to Y$ di due insiemi con strutture di ordinamento (di insiemi ordinati) si dice monotona (rispett. strettamente monotona), se da $x_1 \leq x_2$ ($x_1, x_2 \in X$) segue che $f(x_1) \leq f(x_2)$ (rispettivamente, se da $x_1 < x_2$ segue che $f(x_1) < f(x_2)$). Si può considerare la categoria degli insiemi ordinati con le applicazioni monotone, o strettamente monotone quali morfismi; la seconda è una sottocategoria (incompleta) della prima categoria.

In topologia e nell'algebra omologica gioca un ruolo fondamentale la categoria degli insiemi $\{0, \ldots, n\}$, dove n varia tra tutti i numeri interi non negativi, ordinati nel modo ordinario, con le applicazioni monotone quali morfismi.

Categoria delle varietà differenziabili. Siano X, Y varietà differenziabili. Un'applicazione continua $f : X \to Y$ si chiama morfismo in questa categoria se, localmente, è data da insiemi di applicazioni differenziabili $f : R^m \to R^n$ (cfr. § 3).

Analogamente si definiscono i morfismi di varietà analitiche.

Categoria di fasci su uno spazio topologico X. Siano F, G due fasci di insiemi su uno spazio topologico X. Un morfismo dei fasci $f : F \to G$ è dato da un insieme di applicazioni $f_U : \Gamma(U, F) \to \Gamma(U, G)$, una per ogni insieme aperto $U \subseteq X$. Queste applicazioni devono concordare con le applicazioni-restrizione, che s'indicheranno nello stesso modo, $r_{V,U}$, sia per F che per G, nel senso che $r_{V,U} \circ f_V = f_U \circ r_{V,U}$ per tutti gli U e V, $U \subseteq V$. La composizione di $f : F \to G$ e $g : G \to H$ si definisce mediante un insieme di composizioni $g_U \circ f_U : (U, F) \to \to (U, H)$.

Se F, G sono fasci di gruppi, sui loro morfismi si possono porre ulteriori condizioni affinché tutte le f_U siano morfismi di gruppi. Ciò conduce alla categoria dei fasci di gruppi su X. Analogamente si definiscono le categorie dei fasci di anelli, di spazi lineari, ecc.

In tutti questi esempi la composizione dei morfismi si definisce mediante l'ordinaria composizione teoretico-insiemistica delle composizioni (cfr. § 3), mentre i morfismi identici sono indotti dalle applicazioni identiche dei rispettivi insiemi su se stessi.

SECONDO GRUPPO DI ESEMPI. In questo gruppo di esempi gli oggetti sono, come prima, insiemi con una struttura di tipo definito, mentre i morfismi non sono piú applicazioni di questi insiemi.

Un esempio tipico è la categoria fondamentale della topologia omotopica. I suoi oggetti sono spazi topologici, ma un solo morfismo da X in Y è una classe intera di applicazioni continue $f : X \to Y$, ogni due elementi della quale sono legati tra loro da una relazione di omotopia, cioè possono essere «deformati» uno nell'altro. Un'esatta definizione di omotopia tra le applicazioni $f_0, f_1 : X \to Y$ è la seguente: si tratta di una applicazione continua $F : [0, 1] \times X \to Y$ tale che, se si indica con $F_t : X \to Y$, per $t \in [0, 1]$, l'applicazione $F_t(x) = F(\{t\} \times x)$, allora $F_0 = f_0$ e $F_1 = f_1$. La verifica degli assiomi delle categorie, in particolare l'associatività della composizione dei morfismi (che viene definita quale classe di composizione di rappresentanti arbitrari di questi morfismi) è riportata all'inizio di ogni corso di topologia omotopica.

Un altro esempio è la categoria delle relazioni additive. I suoi oggetti sono tutti i possibili gruppi abeliani. Si chiama morfismo $f : X \to Y$ ogni sottogruppo del prodotto diretto $X \times Y$. La composizione dei morfismi $\varphi : X \to Y$ e $\psi : Y \to Z$ è definita dalla relazione

$$\psi\varphi = \{\langle x, z\rangle \in X \times Z \mid \text{esiste } y \in Y \text{ tale che } \langle x, y\rangle \in \varphi,\ \langle y, z\rangle \in \psi\}.$$

Esistono importanti costruzioni analoghe per altre categorie; per esempio, in geometria algebrica si considera la categoria importante delle varietà algebriche con relazioni quali morfismi.

TERZO GRUPPO DI ESEMPI. Di questo gruppo di esempi fanno parte alcune classiche forme di strutture su un insieme individuale, che, qualche volta, è comodo considerare quali categorie, senza cambiare sostanzialmente la definizione fondamentale.

a) Sia G un certo gruppo. Leghiamo ad esso una categoria, nella quale si ha un unico oggetto e; l'insieme dei morfismi $\mathrm{Hom}(e, e)$ coincide con G e la composizione dei morfismi coincide con il prodotto in G.

b) Sia I un certo insieme ordinato. Leghiamo ad esso una categoria, i cui oggetti sono gli elementi di I, mentre l'insieme dei morfismi $\mathrm{Hom}(x, y)$ è costituito da un solo elemento, se $x \leq y$, ed è vuoto nel caso contrario. La composizione è definita univocamente da queste condizioni. Questa interpretazione è in uso particolarmente quando I è un insieme di indici di un certo sistema induttivo, diciamo, di gruppi.

c) Sia E un certo spazio topologico. Leghiamo ad esso una categoria, i cui oggetti sono i sottoinsiemi aperti di E, mentre $\mathrm{Hom}(x, y)$ consiste di un solo elemento, se $x \subseteq y$, ed è vuoto nel caso contrario.

Questa riformulazione banale costituí il punto di partenza per una profonda generalizzazione della nozione di spazio topologico; si tratta delle cosiddette «topologie di Grothendieck».

d) Consideriamo un qualsiasi grafo orientato $\Gamma = (I, F, d : F \to I \times I)$. Qui I è l'insieme dei vertici del grafo, F è l'insieme dei suoi archi, mentre l'applicazione d associa ad ogni arco una coppia ordinata di vertici, cioè l'origine e la fine di quest'arco.

Al grafo Γ si legano in modo naturale due diverse categorie, i cui oggetti sono gli elementi di I.

In una di esse $\mathrm{Hom}(x, y)$ è costituito da tutti i possibili percorsi nel grafo Γ da x a y lungo gli archi orientati.

Nell'altra categoria $\mathrm{Hom}(x, y)$ contiene un solo elemento, se esiste un percorso da x a y lungo gli archi, ed è vuoto nel caso contrario.

Queste categorie si possono chiamare categorie, rispettivamente, di schema di diagrammi di tipo Γ e di schema di diagrammi commutativi di tipo Γ. Il significato di queste denominazioni lo chiariremo piú oltre.

Esiste una serie di utili costruzioni che permettono, a partire da date categorie, di costruirne di nuove.

Categoria duale. Sia C una certa categoria. La categoria C^0 ad essa duale è data cosí: $ObC^0 = ObC$, $\mathrm{Hom}_{C^0}(X, Y) = \mathrm{Hom}_C(Y, X)$. In altre parole, C^0 si ottiene da C «invertendo tutte le frecce».

Questa costruzione risulta utile in due casi estremi. Se la categoria C^0 è vicina alla C (p. es., se è equivalente ad essa, come nel caso della categoria dei gruppi abeliani finiti), questo stesso fatto è l'espressione d'una certa «legge di dualità».

Al contrario, per la categoria $\mathcal{R}$ degli anelli commutativi la sua categoria duale «degli schemi affini» è molto lontana da $\mathcal{R}$; l'inversione delle frecce fa rilevare delle proprietà geometriche inaspettate; permette di definire l'incollatura di oggetti globali a partire da quelli locali e altre importanti operazioni «geometriche», al limite innaturali dal punto di vista di $\mathcal{R}$.

Categoria di oggetti sopra una data base. Sia C una certa categoria, S sia un suo oggetto prefissato. Introduciamo la categoria C_S con le seguenti definizioni.

ObC_S = tutti i possibili morfismi $X \xrightarrow{\varphi} S$ della categoria C. Se $\varphi : X \to S$, $\psi : Y \to S$ sono due di tali morfismi, $\mathrm{Hom}_{C_S}(\varphi, \psi)$ è definito quale insieme dei morfismi $\chi : X \to Y$ per i quali $\psi \circ \chi = \varphi$. La composizione dei morfismi in C_S è indotta dalle composizioni in C. La correttezza delle definizioni e gli assiomi delle categorie si verificano per C_S senza difficoltà. Invertendo le frecce in queste definizioni, si giunge alla costruzione duale della categoria C^S degli oggetti di C «sopra» l'oggetto S.

Queste costruzioni generalizzano tali nozioni, quali quella di K-algebre (cioè di un anello A insieme all'omomorfismo strutturale di anelli $K \to A$), oppure quella di «fibrato generalizzato» sopra una base B, cioè di uno spazio topologico E insieme con le applicazioni continue $E \to B$.

Prodotto di categorie. Siano C_1, C_2 due categorie. Definiamo il prodotto cartesiano $C = C_1 \times C_2$ ponendo $Ob(C_1 \times C_2) = ObC_1 \times ObC_2$. Qualunque siano gli oggetti $X = \langle X_1, X_2 \rangle$, $Y = \langle Y_1, Y_2 \rangle$ di C, dove $X_1, Y_1 \in ObC_1$; $X_2, Y_2 \in ObC_2$, poniamo $\mathrm{Hom}_C(X, Y) = \mathrm{Hom}_{C_1}(X_1, Y_1) \times \mathrm{Hom}_{C_2}(X_2, Y_2)$. La composizione dei morfismi in C è definita componente per componente.

I gruppi nelle categorie. La definizione ordinaria teoretico-insiemistica di un gruppo X consiste nel dare una legge di composizione sull'insieme degli elementi di X che soddisfa determinati assiomi. Mostriamo che questa definizione si può formulare in termini puramente categoriali e può essere estesa ad oggetti di categorie più generali.

a) Legge di composizione su X. Si tratta di un'applicazione $X \times X \xrightarrow{m} X$. In ogni categoria «con i prodotti diretti (cartesiani)» essa può essere sostituita assegnando il morfismo corrispondente. Il prodotto diretto degli oggetti X, Y nella categoria C è definito quale oggetto Z insieme con i morfismi $Z \xrightarrow{p_1} X$ e $Z \xrightarrow{p_2} Y$ e tale che per ogni oggetto U si ha un isomorfismo degli insiemi $\mathrm{Hom}_C(U, Z) \to \mathrm{Hom}_C(U, X) \times \mathrm{Hom}_C(U, Y)$ che associa a ogni morfismo $U \xrightarrow{f} Z$ la coppia di morfismi $U \xrightarrow{p_1 \circ f} X$ e $U \xrightarrow{p_2 \circ f} Y$. Una categoria C si dice categoria con prodotti diretti se per ogni due oggetti esiste in essa il prodotto diretto (due qualsiasi prodotti diretti di X e Y sono canonicamente isomorfi).

b) Inversione. Si tratta di un'applicazione $i : X \to X$. Nella definizione teoretico-insiemistica si ha $i(x) = x^{-1}$ per ogni $x \in X$. In ogni categoria possiamo assegnare il morfismo corrispondente.

c) L'unità di un gruppo. Invece d'indicare un elemento distinto dell'insieme X, possiamo assegnare il morfismo degli insiemi $\{\emptyset\} \to X$, dove $\{\emptyset\}$ è l'insieme standard con un solo elemento nell'universo di Zermelo-Fraenkel. Per definire l'analogo di $\{\emptyset\}$ in una categoria qualsiasi osserviamo che l'insieme delle applicazioni di un insieme arbitrario X in $\{\emptyset\}$ consiste di un unico elemento. Assumiamo ciò per caratterizzare un oggetto finito E di ogni categoria C: un oggetto E si dice finito, se $\mathrm{Hom}(X, E)$ consiste di un solo elemento per ogni $X \in ObC$. Dopo di che l'unità di un oggetto gruppale X è assegnata quale morfismo $E \to X$ in C.

d) Assiomi di un oggetto gruppale. Un oggetto X di una categoria C insieme con i morfismi $m : X \times X \to X$, $i : X \to X$ ed $e : E \to X$ si chiama oggetto gruppale se soddisfa i seguenti tre assiomi:

1) associatività: le composizioni dei morfismi $X \times X \times X \xrightarrow{(m, \mathrm{id}_X)} X \times X \xrightarrow{m} X$ e $X \times X \times X \xrightarrow{(\mathrm{id}_X, m)} X \times X \xrightarrow{m} X$ coincidono;
2) inverso sinistro: le composizioni dei morfismi $X \xrightarrow{\delta} X \times X \xrightarrow{(i, \mathrm{id}_X)} X \times X \xrightarrow{m} X$ e $X \xrightarrow{\cdot} E \xrightarrow{e} X$ coincidono. Qui $X \xrightarrow{\delta} X \times X$ è il morfismo diagonale, cioè tale che le composizioni di $X \xrightarrow{\delta} X \times X$ con $X \times X \xrightarrow{p_1, p_2} X$ sono uguali a id_X, mentre $X \xrightarrow{\cdot} E$ è l'unico morfismo esistente;
3) unità sinistra: la composizione dei morfismi $X \xrightarrow{\delta} X \times X \xrightarrow{(\cdot, \mathrm{id}_X)} E \times X \xrightarrow{(e, \mathrm{id}_X)} X \times X \xrightarrow{m} X$ coincide con id_X.

Si può definire la nozione di morfismo di due oggetti gruppali nella categoria C e quindi la nozione di «categoria di oggetti gruppali» della categoria C. Se, quale C, si prende l'universo teorico-insiemistico, giungiamo all'ordinaria categoria dei gruppi.

Se, quale C, si prende la categoria degli spazi topologici, giungiamo alla categoria dei gruppi topologici.

Un esempio importante di categoria C, nella quale sono soddisfatte tutte le ipotesi occorrenti per definire gli oggetti gruppali, è la categoria $\mathcal{R}^0$, duale della categoria degli anelli commutativi (con unità), cioè la categoria degli schemi affini. In essa il prodotto diretto degli anelli X, Y è il cosiddetto prodotto tensoriale $X \otimes Y$, mentre l'oggetto finito è l'anello Z dei numeri interi. In tale categoria un oggetto gruppale si chiama «schema affine di gruppi». Se, invece della categoria $\mathcal{R}^0$, si considera la categoria $\mathcal{R}^0_K$, dove K è un certo anello, giungiamo alla nozione di schema affine di K-gruppi.

Funtori.

Anche la nozione di funtore è un caso particolare di applicazione, ma usata nelle categorie. Siano date due categorie C, D. Si chiama funtore F dalla categoria C alla categoria D il seguente insieme di dati:

una applicazione $ObC \xrightarrow{F} ObD$;
una applicazione $F_{X,Y} : \mathrm{Hom}_C(X, Y) \to \mathrm{Hom}_D(F(X), F(Y))$, per ogni X, $Y \in ObC$.

Questi dati devono essere soggetti alla condizione seguente: se $X \xrightarrow{f} Y$, $Y \xrightarrow{g} Z$ sono due morfismi nella categoria C, si ha $F_{Y,Z}(g) F_{X,Y}(f) = F_{X,Z}(gf)$ nella categoria D.

Qualche volta tali F si dicono funtori covarianti dalla categoria C in D, mentre i funtori da C^0 in D si dicono funtori contravarianti da C in D. I funtori contravarianti «invertono le frecce». Spesso i funtori appaiono quali

«costruzioni naturali» sugli oggetti di una categoria C che li trasformano in oggetti di un'altra categoria.

Se C è una categoria arbitraria e Set è la categoria degli insiemi, a partire da ogni oggetto X della categoria C si possono costruire le applicazioni $h_X : Y \to \mathrm{Hom}(X, Y)$ e $h^0_X : Y \to \mathrm{Hom}(Y, X)$. L'applicazione h_X si estende in modo naturale a un funtore da C in Set, mentre h^0_X si estende ad un funtore da C^0 a Set. I funtori di tale forma e cosí pure i funtori ad essi isomorfi, nel significato naturale della parola, si dicono *rappresentabili*. La nozione di funtore rappresentabile divenne una delle prime nozioni fruttuose di carattere categoriale generale dopo che, come risultò, molti funtori importanti, che venivano costruiti in modo del tutto diverso, erano rappresentabili. Per esempio, in topologia ad ogni spazio topologico X di una determinata classe, ad ogni gruppo abeliano G e ad ogni numero intero $n \geq 0$ si associa un invariante importante, cioè il gruppo delle coomologie $H^n(X, G)$. Questo gruppo inizialmente veniva definito mediante alcune costruzioni sopra lo spazio X e sopra G. Risultò che il funtore $X \to H^n(X, G)$ è isomorfo a un certo funtore della forma h^0_Y nella categoria della topologia omotopica, dove $Y = K(G, n)$ è il cosiddetto spazio di Eilenberg-MacLane. Questa scoperta condusse a sviluppare una tecnica completamente nuova in topologia e alla dimostrazione di importanti risultati che erano impossibili con i vecchi mezzi. Come dimostrarono A. Grothendieck e la sua scuola nelle ricerche successive, una gran parte delle questioni di geometria algebrica, per esempio, la classificazione delle diverse varietà algebriche, pure viene ricondotta al problema della rappresentabilità di questo o quel funtore.

Si può dimostrare che un oggetto X è definito univocamente, a meno di isomorfismi, dal funtore h_X, o h^0_X. Ciò permette di considerare ogni categoria quale sottocategoria dei funtori su essa, se preliminarmente è definita la nozione di morfismo dei funtori, o di trasformazione naturale dei funtori. Questo punto di vista conduce ad una estensione naturale di molte classi di oggetti, i cui casi particolari precedentemente venivano considerati senza una esatta comprensione del loro significato categoriale generale.

Infine, molte costruzioni teoretico-strutturali possono essere considerate quali funtori, ciò che conduce a una migliore comprensione delle loro proprietà. Per esempio, è abbastanza nota la nozione di diagramma di strutture di un dato tipo, connesse alle applicazioni: cosí la scrittura $X \to Y$ è un diagramma del tipo $\cdot \to \cdot$; la scrittura

$$\begin{tikzcd} X \arrow[r] \arrow[dr] & Y \arrow[d] \\ & Z \end{tikzcd}$$

è un diagramma del tipo

$$\begin{tikzcd} \cdot \arrow[r] \arrow[dr] & \cdot \arrow[d] \\ & \cdot \end{tikzcd}$$

Tutti i diagrammi di un dato tipo Γ possono essere considerati quale funtore da una categoria di schema di diagrammi di tipo Γ alla categoria dove giac-

ciono X, Y, Z, ... (cfr. sopra). Gli schemi di diagrammi commutativi di tipo Γ conducono a diagrammi commutativi quando le composizioni di tutti i morfismi del diagramma lungo ogni percorso da un oggetto ad un altro coincidono.

Nella matematica moderna l'importanza delle concezioni categoriali nello studio delle applicazioni è molto grande e sarà ulteriormente chiarita in altri articoli. [JU. I. M.].

Cantor, G.
1879-83 *Über unendliche, lineare Punktmannigfaltigkeiten*, in «Mathematische Annalen», XV (1879), pp. 1-7, XVII (1880), pp. 355-58, XX (1882), pp. 113-21, XXI (1883), pp. 51-58 e pp. 545-91.

Cohen, P. J.
[1963-64] *The Independence of the Continuum Hypothesis*, in *Set Theory and the Continuum Hypothesis*, W. A. Benjamin, New York 1966.

Dedekind, R.
1888 *Was sind und was sollen die Zahlen?*, Vieweg, Braunschweig (trad. it. in *Essenza e significato dei numeri. Continuità e numeri irrazionali*, Stock, Roma 1926, pp. 7-118).

Descartes, R.
1637 *Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences (Géometrie)*, Maire, Leyda.

Heyting, A.
1956 *Intuitionism. An Introduction*, North-Holland, Amsterdam.

Milnor, J. W.
1956 *On manifolds homeomorphic to the seven-sphere*, in «Annals of Mathematics», LXIV, pp. 399-405.

Riemann, B.
[1851] *Grundlagen für eine allgemeine Theorie der Functionen einer veränderlichen complexen Grösse* (tesi di laurea, Göttingen) (trad. it. in «Annali di matematica pura e applicata», II (1859)).

Shannon, C.
1948 *A mathematical theory of communication*, in «Bell System Technical Journal», XXVII, 3, pp. 379-423, e 4, pp. 623-56.

Thom, R.
1972 *Stabilité structurelle et morphogénèse. Essai d'une théorie générale des modèles*, W. A. Benjamin, New York.

Whitehead, A. N., e Russell, B.
1962 *Principia Mathematica to *56*, Cambridge University Press, London.

Il concetto di applicazione tra insiemi astratti (cfr. **insieme**) è il frutto del progressivo astrarsi e generalizzarsi del concetto di funzione (cfr. **funzioni**) all'interno del **calcolo** differenziale (cfr. **differenziale, infinitesimale**). Con l'aiuto del concetto astratto di applicazione è possibile ridiscendere all'interno di **strutture matematiche** particolari esaminando, nello specifico, situazioni ove intervengono applicazioni concrete (cfr. **astratto/concreto**). Ciò conduce a varie particolarizzazioni e suddivisioni all'interno del concetto (cfr. **locale/globale**, **geometria/topologia**, **continuo/discreto**). Dal punto di vista della calcolabilità effettiva le applicazioni pongono problemi teorici e pratici di grande rilievo (cfr. **ricorsività, algoritmo, automa**) che costringono continuamente a ridiscutere il rapporto tra matematica «speculativa» (cfr. **matematiche**) e **scienza** applicata.

Assioma/postulato

1. *Generalità.*

Scopo di ogni scienza naturale è la conoscenza della verità sulla Natura e sull'Uomo. Di solito una scienza inizia con una parte descrittiva, volta a delineare l'arco dei fenomeni studiati e a inventariare le nozioni fondamentali della futura teoria. A questo livello viene introdotto anche il linguaggio della teoria e si stabiliscono le regole per confrontare i testi [proposizioni], in questo linguaggio, con la realtà da essi descritta, cioè con la semantica del linguaggio. Possono essere esempi di scienze descrittive la sistematica in biologia, o le grammatiche descrittive delle lingue umane.

Il livello successivo consiste nell'individuare le leggi fondamentali alle quali è soggetta la struttura degli oggetti e dei fenomeni studiati, il loro sviluppo nel tempo e il loro comportamento in determinate condizioni. La certezza sulla veridicità delle leggi è fondata su una induzione incompleta che parte da osservazioni, esperimenti e dati indiretti, il volume dei quali non può mai essere completo. In sostanza è da questo momento che comincia una conoscenza teorica che permetta di predire tratti caratteristici di situazioni ancor prima che esse siano direttamente osservabili nella realtà e senza queste stesse osservazioni dirette.

A misura del suo sviluppo la teoria incorpora in sé frammenti di carattere puramente deduttivo sempre piú lunghi. In questi frammenti gli enunciati, nuovi e veri, si ottengono dalle verità sperimentali e dalle leggi della teoria mediante una serie di manipolazioni logiche, dette inferenze deduttive, dimostrazioni, ecc. La parte deduttiva della teoria si sviluppa sulla base delle informazioni che già si hanno, senza l'uso di conoscenze supplementari. Quanto piú lunghi sono i frammenti deduttivi della teoria, tanto piú pressante diventa la necessità di individuare e inventariare le premesse e la logica dei ragionamenti deduttivi che vengono usati.

Questo schema è applicabile anche alla matematica, ma soltanto con delle convenzioni essenziali.

Rispetto agli altri campi del sapere la matematica si distingue per una elevata astrazione dei suoi concetti fondamentali e per il ruolo dominante che in essa hanno i ragionamenti deduttivi.

L'astrazione dei concetti fondamentali si esprime nel fatto che il linguaggio della matematica e le sue sottoteorie ammettono una enorme varietà di semantiche, o di interpretazioni del tutto diverse tra loro. Ciò dà la possibilità, per esempio, ai fisici di considerare la stessa matematica quale linguaggio delle teorie fisiche. Questa astrazione già diventa evidente considerando le nozioni matematiche piú elementari, quali, per esempio, i semplici numeri naturali (uno, due, tre). Tali nozioni erano già note alle popolazioni che si trovavano ad uno stadio di sviluppo assai primitivo. Secondo gli ultimi dati della linguistica comparata, nella lingua con cui si parlava ancora nel Mesolitico (piú di diecimila anni fa) e che è l'antenata comune a molte famiglie di lingue moderne quali quelle indoeuropee, semito-camitiche, degli Urali, dell'Altaj ed altre, vi era la parola **tohä*, che significa 'due' (il cui aspetto sonoro è giunto fino ai nostri giorni con piccole variazioni nelle diverse lingue). Per rendersi conto della vastità della nozione 'due' al lettore basterà confrontare le sue interpretazioni nei contesti che seguono: «due pesci», «due giorni», «dualismo filosofico» (dal latino *duo* 'due'), «la sesta cifra dopo la virgola nella scrittura decimale di π è 2».

La struttura deduttiva delle teorie matematiche fornisce dei modelli di costruzioni logiche cosí sviluppati, complessi e lunghi, che essa stessa, la matematica, costituisce un modello universalmente riconosciuto per costruzioni teoriche in ogni dominio del sapere. In misura notevole ciò è legato al fatto che l'apparato evoluto del simbolismo matematico costituisce un esempio di linguaggio artificiale libero dagli equivoci e dalle limitazioni celati nelle lingue correnti. Il linguaggio dei simboli, non avendo un legame diretto con la realtà, permette di descrivere una infinita varietà di situazioni e sviluppare delle teorie che, a priori, non sono collegate a questo o a quel campo di fenomeni.

Il concetto di assioma appartiene a quel numero di nozioni fondamentali che fanno parte della metateoria delle teorie deduttive (cfr. § 3). La parola 'assioma', dal greco ἀξίωμα 'dignità', indica una di quelle premesse di una data teoria deduttiva, la cui veridicità viene assunta, internamente a questa teoria, senza dimostrazione e viene poi usata nel seguito come un tutto unico. È possibile produrre un'argomentazione a favore della veridicità di un assioma (o, quanto meno, a favore di una sua assunzione provvisoria). Tuttavia, rigorosamente parlando, essa non appartiene alla teoria. La questione riguardante la natura della verità matematica, in generale, è molto complessa e, piú oltre, al § 5, sarà discussa piú particolareggiatamente.

La parola 'postulato', dal latino *postulatum* 'si richiede', si usa attualmente come sinonimo di assioma. Tuttavia vi sono sfumature diverse nell'uso delle due parole che non sono completamente normalizzate. Negli *Elementi* di Euclide, secondo Proclo, i postulati affermano l'esistenza di costruzioni di oggetti geometrici, mentre gli assiomi descrivono le proprietà di questi oggetti. Alcuni specialisti contemporanei di logica matematica chiamano postulati alcuni presupposti speciali di una data teoria particolare, a differenza degli assiomi logici e generali. Infine si possono chiamare postulati quei presupposti di una data teoria assunti, magari temporaneamente, per ricavare da essi delle conseguenze, confrontando le quali con questa o quella realtà, si può giudicare dei limiti nell'uso di questi stessi postulati. I fisici parlano, per esempio, dei «postulati della teoria del campo».

A scopo illustrativo vediamo alcuni esempi di assiomi senza limitarci a quelli strettamente matematici. Al § 4 verrà esposto un modello completo del sistema di assiomi della teoria degli insiemi, che è alla base della matematica moderna.

ASSIOMA DELL'UGUAGLIANZA. *Due oggetti, ognuno dei quali sia uguale ad un terzo, sono uguali tra loro.*

Questo assioma è già presente negli *Elementi* di Euclide. Nelle costruzioni matematiche teoretico-insiemistiche moderne esso interviene quale proprietà fondamentale delle relazioni di equivalenza: la transitività di queste relazioni.

SILLOGISMI DI ARISTOTELE. Ecco con linguaggio moderno uno degli schemi dei sillogismi: *« Se tutti gli oggetti di una classe A godono di una proprietà P e a è un oggetto della classe A, a gode della proprietà P »* (« Tutti gli uomini sono mortali, Caio è un uomo, quindi Caio è mortale »).

Gli assiomi dell'uguaglianza e i sillogismi appartengono agli assiomi logici; la loro verità viene postulata indipendentemente dalla natura e dalle proprietà degli oggetti considerati, cioè indipendentemente dall'interpretazione della teoria. In essi viene distintamente seguito l'aspetto linguistico-semantico; essi regolano l'uso corretto delle parole « essere uguali », « tutti » e simili. Nelle teorie formalizzate l'aspetto linguistico interviene in primo piano; l'assioma diventa semplicemente un testo unico completamente definito e ogni riferimento ad esso avviene secondo regole sintattiche completamente definite.

Bisogna osservare che l'uso contenutistico persino di un assioma fondamentale qual è l'assioma dell'uguaglianza può richiedere che siano individuati aspetti non banali della situazione descritta. Per esempio, nella fisica quantistica si parla della indistinguibilità in linea di principio di tutte le particelle elementari di una data classe, – per esempio, degli elettroni, – ammettendo contemporaneamente (nella teoria non relativistica) che, negli intervalli di tempo fra una interazione e l'altra, l'elettrone mantenga la sua individualità.

Gli assiomi seguenti appartengono a parti diverse della matematica propriamente detta.

QUINTO POSTULATO DI EUCLIDE (ASSIOMA DELLE PARALLELE). *Se due rette intersecate da una terza retta formano una coppia di angoli coniugati interni la cui somma è minore di 180°, queste due rette si intersecano in un punto che giace, rispetto alla terza retta, dalla stessa parte dei due angoli coniugati.*

Questo assioma ha avuto un ruolo notevole nella formazione della matematica moderna e dei suoi fondamenti. Per molti secoli la sua verità è apparsa meno evidente che non quella degli altri assiomi di Euclide. Per questo sono stati intrapresi molti tentativi per dedurlo dai restanti assiomi di Euclide. Nei lavori di Lobačevskij, Bolyai e Gauss fu costruita una teoria di geometrie non euclidee, quale insieme di deduzioni ricavate dalla negazione del quinto postulato e dai rimanenti assiomi di Euclide. Questa teoria si spinse assai avanti senza incontrare contraddizioni, tanto che sorse la convinzione psicologica della non-deducibilità del quinto postulato dai restanti assiomi e quindi della possibile sua falsità nel mondo fisico (nella interpretazione, per esempio, delle rette euclidee quali raggi luminosi, o delle traiettorie del moto delle particelle libere). Tuttavia, il fatto che fino a un dato momento non si sia scoperta nella teoria una contraddizione non significa ancora che non si scoprirà mai. Evidentemente Gauss era chiaramente consapevole di questa circostanza, ciò che probabilmente l'ha indotto a rinunziare, in vita, alla pubblicazione delle

sue ricerche sulla geometria non-euclidea. La dimostrazione del fatto che il quinto postulato di Euclide non sia effettivamente deducibile dai rimanenti assiomi si è accompagnata all'introduzione di un nuovo principio, destinato a diventare un metodo importante nella logica matematica del XX secolo. Si tratta del principio della interpretazione interna non-standard di una teoria matematica, o metodo dei modelli. Nell'uso di questo metodo si propone una nuova concezione (internamente alla matematica) di tutti i termini della teoria matematica, per la quale tutti gli assiomi, ad eccezione di quello la cui non-deducibilità si vuole verificare, risultino veri, e quest'ultimo, invece, risulti falso. Per esempio, esiste un modello di geometria non-euclidea, nel quale al piano euclideo corrisponde un semipiano, ai punti euclidei corrispondono i punti del semipiano, mentre alle rette euclidee corrispondono semicirconferenze appoggiate sulla frontiera del semipiano. Tutte le altre nozioni primitive vengono reinterpretate in modo corrispondente. In questo modello il quinto postulato è falso.

ASSIOMA DELL'INDUZIONE DI PEANO. *Sia P una certa proprietà dei numeri naturali. Supponiamo che* 1 *goda della proprietà P e che, per ogni n, dal fatto che n goda della proprietà P segua che* $n+1$ *goda della stessa proprietà P. Allora tutti i numeri naturali godono della proprietà P.*

L'assioma dell'induzione è il primo principio matematico fondamentale in cui si fa riferimento all'infinito. Infatti esso si appella alla possibilità astratta di produrre una serie infinita di verifiche, quanto si voglia lunghe, per convincerci sulla validità di P per ogni n, muovendo a passo a passo da 1 a 2, da 2 a 3, da 3 a 4, ..., da $n-1$ a n. È pressoché l'unico « principio dell'infinito » che si incontra nelle ricerche metamatematiche, dove di regola, nei sistemi formali, le affermazioni contenute nei testi sono stabilite per induzione secondo la lunghezza dei testi medesimi. Internamente alla stessa matematica « la teoria elementare dei numeri » incorpora dei corollari di questo assioma e un numero limitato di principî « finiti ». Infine Georg Cantor [1895-97] nella sua teoria degli insiemi ha suggerito una notevole generalizzazione di questo assioma, « il principio dell'induzione transfinita », che, in forme diverse equivalenti, è diventato uno degli strumenti piú potenti, ma anche piú contestati, della matematica del XX secolo.

ASSIOMA DELL'ASSOCIATIVITÀ. *Sia M un certo insieme e supponiamo che per ogni due elementi, x, y, di M sia definito un terzo elemento, xy, di M, detto il loro prodotto. Se per tutti gli x, y, z di M è soddisfatta l'uguaglianza* $(xy)z = x(yz)$, *si dice che per il prodotto in M vale l'assioma dell'associatività.*

Piú generalmente le proprietà di una struttura (teorico-insiemistica), postulate nella definizione della stessa struttura, si dicono spesso assiomi di questa struttura.

A conclusione riportiamo due assiomi delle teorie fisiche.

ASSIOMA DELL'ADDITIVITÀ DELLE MASSE. *La massa dell'unione di due corpi distinti è uguale alla somma delle masse di questi corpi.*

ASSIOMA DELLO STATO NELLA TEORIA QUANTISTICA RELATIVISTICA. *Lo stato di un sistema è descritto mediante i raggi unitari in uno spazio di Hilbert separabile. La legge relativistica di trasformazione dello stato è data mediante una rappresentazione unitaria continua del gruppo non omogeneo SL*(2, **C**) [Streater e Wightman 1964].

L'ultimo esempio rende particolarmente chiara la necessità di descrivere preliminarmente il linguaggio della teoria, la sua semantica, e la necessità di una vasta quantità di argomentazioni a favore della verità di un assioma. L'assioma dello stato, che interviene quale presupposto di una nuova teoria, rientra nel novero delle piú ardite generalizzazioni della fisica quantistica classica.

2. *Formazione del metodo assiomatico in matematica.*

Il metodo assiomatico è quel modo di costruire o di esporre una teoria, per il quale le premesse iniziali vengono formulate esplicitamente quali assiomi, mentre tutti i risultati successivi sono ottenuti da questi assiomi a conclusione di successioni di ragionamenti soggetti alle norme della logica.

Il metodo assiomatico è per eccellenza un metodo della matematica. Come si è già notato, le generalizzazioni scientifiche naturali, tipo teoria dell'evoluzione in biologia, o linguistica comparata, si distinguono per una vastità di dati iniziali osservati e per risultati ottenuti con procedimenti deduttivi relativamente brevi. Di piú, la maggior parte di questi risultati si presta ad una loro verifica, ricorrendo, ove occorra anche piú volte, ai dati osservati, di modo che l'esigenza di un loro fondamento logico può essere notevolmente attenuata.

Al contrario, la matematica e frammenti significativi della fisica si distinguono appunto per il fatto che i risultati essenziali sono ottenuti mediante lunghi ragionamenti deduttivi da un numero limitato di premesse. La verifica sperimentale di molti risultati intermedi e di quelli finali risulta persino impossibile per questioni sia tecniche che di principio. Del resto, nella maggioranza dei casi, per le stesse caratteristiche della matematica, non ha neppure senso porre la questione di una tale verifica: l'affermazione, secondo la quale tutti i numeri interi godono di una proprietà P, a rigore, dovrebbe essere il risultato di una successione infinita di verifiche. Solo singoli punti della teoria ammettono un confronto parziale con la realtà in senso pragmatico.

La lunghezza delle inferenze deduttive impone particolari esigenze sulla loro normalizzazione. Per diminuire le probabilità di errore le inferenze sono suddivise in brevi passaggi e, sin dall'inizio, ove sia possibile, si elencano chiaramente i tipi di questi passaggi. L'uso della lingua naturale si caratterizza per il suo impoverimento (per diminuire le omonimie) e per una elevata riserva terminologica.

I primi modelli di teorie assiomatiche, giunti a noi nei loro testi, risalgono al V secolo a. C. È tradizione ascriverli alla scuola pitagorica. Euclide, Archimede e Apollonio ci hanno lasciato modelli di esposizioni matematiche dalle

quali risulta che il concetto di dimostrazione matematica che essi avevano coincide sostanzialmente con quello che si ha attualmente.

Lo sviluppo ulteriore del metodo assiomatico non è stato lineare. Da un lato gli studiosi si rendevano conto dei principî logici dei ragionamenti regolari e una parte integrante dei loro studi consisteva nell'individuare tali principî. Questa linea parte da Aristotele e conduce alla logica formale della scuola di Hilbert del XX secolo. Dall'altra parte, particolarmente a partire dal secolo XVII, si è avuto un aumento notevole del materiale matematico, costituito da nuove idee e fatti diversi. In certi periodi l'attenzione alla chiarezza dei principî fondamentali si indeboliva e ciò per effetto delle enormi possibilità che si aprivano. Nei periodi «critici» successivi si aveva un riesame di tutto il materiale accumulato e, dall'accurata analisi logica cui veniva sottoposto, si individuavano nuove nozioni fondamentali e si chiarivano i legami logici.

Indichiamo brevemente le tappe di questo sviluppo.

I fondamenti della logica furono posti da Aristotele. Lo scopo fondamentale della sua ricerca gli era perfettamente chiaro: formare un breve elenco di regole tali che a un loro impiego sistematico e conseguente sia riconducibile ogni ragionamento regolare indipendentemente dal contenuto di quest'ultimo e dalla veridicità delle sue premesse iniziali. Aristotele non raggiunse completamente il suo scopo. Prima di tutto egli non individuò i cosiddetti quantificatori, universale ed esistenziale, quali componenti autonome del linguaggio logico, benché egli pure distinguesse le proposizioni «universali» e «particolari». Inoltre, se i sillogismi di Aristotele si interpretano come «tautologie» della logica formale contemporanea, risulta che le componenti atomiche di queste tautologie sono di tipo assai particolare; fondamentalmente «ogni A è B», oppure «qualche A è B».

Le tautologie con un numero arbitrario di componenti atomiche furono profondamente studiate nelle scuole degli stoici e dei megarici. Ad essi era già noto sostanzialmente il teorema fondamentale della logica contemporanea dei predicati secondo il quale, fra tutte le affermazioni universalmente vere, vi è una «base finita» di tali affermazioni, con la quale si possono ottenere tutte le altre mediante successive sostituzioni delle componenti atomiche con enunciati arbitrari. Però questi lavori furono dimenticati fino alla loro riscoperta nel XIX secolo.

In Aristotele e poi ancora per molto tempo lo studio della logica dei ragionamenti deduttivi avvenne, quasi del tutto indipendentemente dalla matematica, su materiale della retorica, della teologia e della giurisprudenza.

Nel secolo XVII, con lo sviluppo del simbolismo algebrico nei lavori di Viète e di Cartesio, si iniziarono dei tentativi volti ad indagare la struttura algebrica della stessa logica. Idee particolarmente profonde appartengono a Leibniz. Ripetutamente egli si accinse ad un progetto di un linguaggio simbolico universale, nel quale i ragionamenti fossero equivalenti a «calcoli», cioè ad operazioni formali sopra i testi, tali da poter essere eseguite in una «macchina logica». Allo stesso Leibniz risale il metodo, divenuto poi famoso dopo le scoperte di Gödel [1931], di ricondurre a calcolo «le operazioni sui testi». Questo

metodo consiste nel numerare con i numeri primi i «testi atomici» e quindi le successioni di questi ultimi numerarle con i prodotti dei numeri primi corrispondenti; dopo di che le operazioni sintattiche si trasformano in operazioni aritmetiche.

La riduzione completa della teoria matematica della dimostrazione alla teoria del calcolo fu ottenuta soltanto nel XX secolo nei lavori di Gödel, Post, Turing, Church, dopo gli studi fondamentali di Boole, Peano, Frege, Russell e Whitehead, e Hilbert.

Parallelamente, e per lungo tempo indipendentemente da questa *comprensione della struttura delle teorie assiomatiche*, si ebbe un *processo di organizzazione assiomatica di tutto il crescente volume di conoscenze matematiche*. Gli storici della matematica osservano una attenuazione dello spirito critico e della pedanteria degli autori antichi al tempo delle grandi scoperte dell'analisi e della geometria analitica. Le nuove idee e gli algoritmi per il calcolo numerico, i loro impieghi nella fisica (teoria della gravitazione di Newton, le serie di Fourier, le equazioni delle onde e della propagazione termica, i metodi variazionali) erano cosí ricchi, mentre le regole per operare sui nuovi formalismi, ricercate a tentoni sulla base dell'analogia e dell'intuizione, erano cosí fruttuose che uno studio scrupoloso sul significato e i limiti di applicabilità delle nozioni introdotte appariva in un primo momento non corrispondente allo spirito dei tempi. Si operava con gli infinitesimi e con i numeri immaginari, non solo senza saper spiegare che cosa significassero, ma spesso senza neppure immaginare quale carattere potesse avere una tale spiegazione. Si cercò di uscirne in due direzioni: per un verso, sviluppando un conseguente formalismo algebrico non contraddittorio come se, internamente ad esso, il calcolo fosse esente dalla necessità di chiarirne il significato, e per l'altro, riconducendo le nuove nozioni a quelle già note, intuitivamente assimilate e, pertanto, apparentemente senza pericolo. Entrambe le direzioni risultarono molto fruttuose. La prima condusse, nel XX secolo, al programma di Hilbert, volto a formare delle concezioni generali dei formalismi matematici, quali giochi con i testi simbolici non richiedenti, in linea di principio, alcuna interpretazione. La seconda, dopo che erano state individuate le nozioni fondamentali dell'analisi (Cauchy, Bolzano, Weierstrass, Dedekind), condusse alla loro riduzione alle proprietà dei numeri reali e poi di quelli interi (Dedekind, Peano, Cantor). Infine Cantor, analizzando la struttura comune di queste nozioni, ha proposto e sviluppato la teoria astratta degli insiemi, che è divenuta il fondamento della matematica moderna e che è stata assiomatizzata nel XX secolo da Zermelo, Fraenkel, Neumann, Gödel, Bernays ed altri.

Parallelamente si attuò un'accurata analisi delle assiomatiche geometriche che ebbe poi un ruolo particolare nella formazione del metodo assiomatico. A questa analisi si accompagnarono, da una parte, la scoperta di una serie di principî della geometria euclidea usati implicitamente negli *Elementi* e, dall'altra, una serie di tentativi per mettere ordine, non solo fra la loro interdipendenza, ma anche nei loro rapporti con il mondo fisico (l'assioma delle parallele venne riguardato quale ipotesi fisica; la questione della sua assunzione, o negazione, doveva essere risolta mediante misurazioni astronomiche, o geodetiche). Abbiamo già accennato alla storia della dimostrazione della non-deducibilità dell'assioma delle parallele.

Dopo i risultati piú conclusivi di Pasch e di Peano uscí il libro di David Hilbert *Grundlagen der Geometrie* [1899], definito da Bourbaki la «carta dell'assiomatica moderna». Hilbert suddivise gli assiomi in alcuni gruppi; esaminò accuratamente quali di essi sono necessari in queste o quelle dimostrazioni e quali sono superflui e costruí uno spettro intero di geometrie differenti, ottenute sostituendo, o tralasciando, assiomi diversi. Inoltre Hilbert stabilí la non-contraddittorietà e la reciproca indipendenza degli assiomi mediante il metodo dei modelli interni, costruendo geometrie multiformi di numeri reali.

Il tentativo piú conseguente e fondamentale di una esposizione assiomatica di tutte le principali concezioni matematiche moderne è stato intrapreso nel XX secolo da un gruppo di matematici francesi, che operano sotto lo pseudonimo collettivo di N. Bourbaki (la sua composizione nel corso di oltre trent'anni è cambiata; attualmente ad esso appartengono anche scienziati di altri paesi). Questa esposizione è l'oggetto di un trattato in piú volumi, *Eléments de mathématique* [1939 sgg.] sottoposto a una regolare rielaborazione e la cui edizione non è ancora completa. L'esposizione è articolata attorno alla nozione centrale di struttura matematica, di cui sono esempi le strutture di legge di composizione, di ordinamento, di topologia, ecc. N. Bourbaki sottolinea il ruolo enormemente euristico di una ben organizzata esposizione assiomatica. Per esempio, quegli oggetti tradizionali della matematica classica che sono i numeri reali, sono un insieme munito delle strutture d'ordine, di corpo, di spazio topologico, di varietà differenziale e di altre. Non appena queste strutture siano state profondamente studiate, ciascuna separatamente, la loro interazione su un oggetto concreto diventa molto piú comprensibile. Soprattutto l'esperienza dello studio e del discernimento delle strutture mette nelle mani di chi opera in matematica un apparato pronto e ricco di nozioni e di metodi per lo studio delle sue questioni specifiche non appena che sia stata ravvisata la presenza di queste strutture nel dominio considerato.

3. *Metateoria delle teorie assiomatiche.*

La metateoria delle teorie assiomatiche sta alle stesse teorie come la linguistica sta alle lingue naturali. La sua formazione nel XX secolo è stata sollecitata sia dalle difficili questioni filosofiche sulla natura della verità in matematica (esse divennero particolarmente acute dopo che furono scoperti i «paradossi» della teoria degli insiemi; cfr. § 5), sia dal desiderio naturale dei matematici di rendersi conto del significato e del carattere della loro attività professionale.

In questo paragrafo ci limiteremo a discutere a un livello qualitativo alcune nozioni matematiche fondamentali. La stessa metateoria può essere in misura notevole formalizzata e assiomatizzata; in questa forma essa appartiene già alla logica matematica.

La metateoria considera ogni teoria assiomatica quale collezione (potenzial-

mente infinita) di testi in un certo linguaggio. Questi testi sono organizzati secondo regole piú o meno esplicitamente formulate che costituiscono la sintassi del linguaggio. La descrizione del linguaggio può essere accompagnata da una sua interpretazione standard, o da uno spettro di tali interpretazioni, ma ciò non è obbligatorio. Rispetto ad una data interpretazione alcuni testi, nel dato linguaggio, possono avere la proprietà di essere veri, o falsi. Le regole d'inferenza di una data teoria sono metodi per generare nuovi testi a partire da altri già dati, o da testi precedentemente generati. Si intende che, in presenza di una interpretazione, le regole d'inferenza applicate a testi veri devono generare di nuovo solo testi veri.

Gli assiomi di una teoria costituiscono un sistema di testi iniziali e questo sistema viene assegnato elencando i testi medesimi, oppure mediante regole per generarli. I testi che si possono ricavare dagli assiomi dati mediante le regole d'inferenza si dicono deducibili.

Nella maggioranza delle teorie vi è il concetto di negazione di un dato testo (enunciato).

Un sistema di testi (in particolare di assiomi) di una teoria si dice non-contraddittorio se da essi non è possibile dedurre un testo e contemporaneamente la sua negazione (cioè una contraddizione).

Un sistema di assiomi si dice completo, rispetto a una data interpretazione e alla nozione di verità, se da questi assiomi si possono dedurre, tutti e soli, i testi veri.

Un sistema di assiomi si dice categorico se esiste solo una interpretazione (in una certa classe), rispetto alla quale tutti gli assiomi risultino veri. Per esempio, il sistema di assiomi di Peano per i numeri interi, in un certo senso (in seno alla teoria degli insiemi), può essere ritenuto categorico, poiché definisce soltanto un oggetto: i numeri interi. Al contrario, il sistema di assiomi della teoria dei gruppi descrive una classe intera di gruppi non isomorfi tra loro.

Un dato testo di un linguaggio si dice indipendente da un sistema di altri testi (per esempio, dal sistema di assiomi) se da questi ultimi non è deducibile né il testo stesso, né la sua negazione. Il quinto postulato di Euclide è indipendente dagli altri assiomi di Euclide.

I problemi fondamentali della metateoria delle teorie assiomatiche sono i seguenti: descrivere il sistema di testi deducibili dagli assiomi; chiarire le questioni relative all'indipendenza, alla completezza, alla non-contraddittorietà, alla categoricità di questo, o di quel sistema di assiomi; stabilire la non-deducibilità o l'indipendenza di questo o di quel testo dagli assiomi; descrivere l'insieme delle interpretazioni, nelle quali i dati assiomi sono veri.

Le nozioni metateoriche sopra riportate sono dotate di una minore definibilità terminologica che non le ordinarie nozioni matematiche. In primo luogo ciò è dovuto al fatto che un'idea del linguaggio della teoria si basa sull'esperienza delle lingue naturali, le quali sono oggetti non formali e molto complessi. L'uso del linguaggio naturale da parte dell'uomo è assai lontano da una manipolazione meccanica dei testi; la generazione della parola, la sua comprensione, il rapporto tra la parola e la realtà, o l'arco delle idee descritte si possono studiare

effettivamente solo ricorrendo ai dati della psicologia, della neurofisiologia, o di altri dominî del sapere, lontani dalla matematica propriamente detta. Ugualmente la logica «naturale» dei ragionamenti deduttivi è, dal punto di vista matematico, una nozione vaga, quantunque i matematici siano relativamente concordi nello stimare logicamente impeccabile, oppure contenente vuoti logici o errori ogni ragionamento dato. Perciò furono intrapresi numerosi tentativi di precisare le definizioni date al punto che la metateoria delle teorie matematiche ha assunto un livello di precisione paragonabile a quello della matematica contemporanea. I risultati di quest'opera sono il contenuto della logica matematica moderna. La metateoria logico-formale studia essenzialmente solo i modelli matematici delle teorie assiomatiche ed è per questo che si ottiene una maggiore precisione nelle sue affermazioni.

Riporteremo piú oltre, nel linguaggio naturale, il sistema di assiomi della teoria degli insiemi secondo Zermelo [1908] e Fraenkel [1928]. Sulla base della teoria degli insiemi, o di suoi frammenti, è costruita quasi tutta la matematica moderna. Per questa ragione gli assiomi della teoria degli insiemi sono «assiomi per eccellenza». Tutti gli altri assiomi matematici, dal punto di vista della teoria degli insiemi, sono soltanto definizioni di oggetti e delle loro proprietà, studiate in questa o in quella particolare teoria come, per esempio, l'assioma dell'associatività del § 1.

Parallelamente alla formulazione degli assiomi introdurremo le stesse nozioni matematiche e le notazioni di uso corrente.

4. *Assiomi della teoria degli insiemi secondo Zermelo-Fraenkel.*

La teoria degli insiemi che descriveremo ha lo scopo di studiare un'unica classe di oggetti fondamentali, detti *insiemi*, ed un'unica relazione binaria fra gli insiemi: «essere elemento». La scrittura $x \in y$ si legge «l'insieme x è un elemento dell'insieme y». Inoltre dagli assiomi di Zermelo-Fraenkel segue che non esiste una catena $x_1 \ni x_2 \ni x_3 \ni \ldots$ di lunghezza infinita. Ne segue che l'ultimo elemento di una catena di questa forma, che non sia prolungabile, è un insieme che non ha elementi. Esso si dice *vuoto* e si indica con $\varnothing$. Un insieme finito, tutti gli elementi del quale siano $x_1, \ldots, x_n$, può essere indicato con $\{x_1, \ldots, x_n\}$. Pertanto tra gli insiemi dell'universo di Zermelo-Fraenkel sono contenuti, per esempio, $\varnothing$, $\{\varnothing\}$, $\{\{\varnothing\},\varnothing\}$, ..., ecc. Una descrizione piú esatta del processo di «collezionamento» di ogni insieme di questo universo a partire dall'insieme vuoto e da insiemi precedentemente costruiti è contenuta negli assiomi che ci accingiamo ad esporre.

ZF1. ASSIOMA DELL'ESTENSIONE. *Due insiemi, x, y, sono uguali se, e solo se, ogni elemento dell'insieme x è un elemento dell'insieme y e ogni elemento dell'insieme y è un elemento dell'insieme x.*

Da un punto di vista formale questo assioma regola l'uso del predicato «essere uguali» riferito agli insiemi. Rispetto al contenuto questo assioma esprime

l'idea fondamentale di Cantor: immaginare un aggregato di oggetti come un tutto unico, definito soltanto dai suoi elementi e nient'altro di piú.

Se gli insiemi x, y sono uguali, scriviamo $x = y$. Se ogni elemento dell'insieme x è un elemento dell'insieme y, scriviamo $x \subseteq y$ e leggiamo «x è un sottoinsieme di y».

ZF2. ASSIOMA DELL'INSIEME VUOTO. *Esiste un insieme che non contiene elementi.*

Dall'assioma dell'estensione segue che l'insieme vuoto è unico, cioè che due qualsiasi insiemi vuoti sono uguali.

Nella verifica di questa affermazione il lettore può porre mente ad una logica indeterminatezza del linguaggio naturale: è lecito affermare che ogni elemento dell'insieme x è un elemento dell'insieme y, quando x non contiene neppure un elemento? Individuando e normalizzando i principî logici utilizzabili, la risposta a questa domanda risulta affermativa.

Questo assioma e tutti i successivi postulano l'esistenza di insiemi con queste o quelle proprietà.

ZF3. ASSIOMA DELLA COPPIA. *Per ogni due insiemi, x, y, esiste un terzo insieme, z, elementi del quale sono soltanto x e y.*

Questo insieme, d'accordo con le nostre convenzioni, si può indicare con $\{x, y\}$; esso è definito univocamente in virtú dell'assioma dell'estensione e si chiama coppia non ordinata degli insiemi x e y. Evidentemente $\{x, x\} = \{x\}$.

Si chiama coppia ordinata degli insiemi x, y, con primo elemento x e secondo y, l'insieme $\{\{x\}, \{x, y\}\}$. Brevemente si indica con $\langle x, y\rangle$. Non è difficile verificare che $\langle x, y\rangle = \langle u, v\rangle$ se, e solo se, $x = u$ e $y = v$. Le coppie ordinate sono usate essenzialmente per definire le nozioni di corrispondenza fra insiemi e di applicazione di insiemi (cfr. oltre).

ZF4. ASSIOMA DELLA SOMMA. *Per ogni insieme x esiste un insieme y tale che l'insieme z è un elemento di y se, e solo se, esso è un elemento di un certo elemento dell'insieme x.*

L'insieme y si dice insieme-somma, o unione di x e si indica con $\bigcup x$. Al posto di $\bigcup\{x_1, x_2\}$ spesso si scrive $x_1 \cup x_2$; al posto di $\bigcup\{x_1, \ldots, x_n\}$ si scrive $\bigcup_{i=1}^{n} x_i$, ecc.

ZF5. ASSIOMA DEL GRADO. *Per ogni insieme x esiste un insieme y tale che gli elementi di y sono, tutti e soli, i sottoinsiemi di x.*

Notazione abbreviata: $y = \mathcal{P}(x)$.

Ricordiamo che z è un sottoinsieme di x, $z \subseteq x$, se ogni elemento di z è un elemento di x. Fra i sottoinsiemi di x vi sono sempre $\varnothing$ e x; gli altri sottoinsiemi di x si dicono *propri*.

Gli assiomi fin qui formulati sono soddisfatti nell'universo costituito da tutti gli insiemi finiti che si ottengono a partire da $\varnothing$ mediante il processo

di «collezionamento» successivo degli insiemi precedentemente già costruiti: $\varnothing, \{\varnothing\}, \{\varnothing,\{\varnothing\}\}, \{\varnothing,\{\varnothing\},\{\varnothing,\{\varnothing\}\}\}$, e cosí via. Per uscire dai limiti di questo universo è necessario l'assioma seguente.

ZF6. ASSIOMA DELL'INFINITO. *Definiamo la nozione di insieme x' seguente all'insieme x: $x' = x \cup \{x\}$. Esiste un insieme x contenente $\varnothing$ e, insieme ad ogni $y \in x$, contiene y'.*

Pertanto questo assioma postula l'esistenza di un insieme contenente tutti i $\varnothing, \varnothing', \varnothing'', \varnothing''', \ldots$

ZF7. ASSIOMA DEL SOTTOINSIEME. *Per ogni insieme x e per ogni proprietà P di un insieme esiste un insieme y tale che $z \in y$ se, e solo se, $z \in x$ e z gode della proprietà P.*

Si usa scrivere $y = \{z | z \in x$ e z gode della proprietà $P\}$.

Nella formulazione di questo assioma per la prima volta appare la nozione generale «proprietà di un insieme» che noi non abbiamo definito. Qui ci troviamo di fronte essenzialmente a un problema linguistico; dobbiamo chiarire quali frasi del linguaggio naturale, contenenti il nome insieme x, siano da ritenere una descrizione delle proprietà dell'insieme x. Una risposta precisa può essere data solo dopo aver introdotto il linguaggio formale della teoria degli insiemi. Noi ci limitiamo a far notare che queste frasi devono comprendere solo le notazioni di insieme, le parole «non», «e», «oppure», «se, ..., allora...», «se, e solo se», «esiste», «per ogni», «uguale», «essere elemento» e cosí pure i loro sinonimi e le parole ausiliarie del linguaggio naturale, e, infine, esse devono essere «costruite regolarmente». La difficoltà maggiore è chiarire cosa significa «costruzione regolare» di una frase. Farlo nei limiti del linguaggio naturale è oltremodo scomodo e questa è una delle ragioni principali che giustificano l'introduzione dei linguaggi formali.

Sulle proprietà esiste un altro punto di vista complementare. Ogni proprietà di certi elementi, diciamo, di un insieme x, definisce in x, come si può mostrare, il sottoinsieme costituito dagli elementi che godono di questa proprietà. Inversamente, ogni sottoinsieme $y \subseteq x$ definisce la proprietà «essere elemento di y». Perciò, d'accordo con questo punto di vista «estensionale», le proprietà degli elementi di x possono essere identificate con i sottoinsiemi di x. Però, se x è infinito, per esempio, se contiene $\varnothing, \varnothing', \varnothing'', \ldots$, ne segue che tutti i sottoinsiemi di x risultano notevolmente di piú che non tutti i sottoinsiemi che possono essere esplicitamente descritti con frasi del linguaggio regolarmente costruite. In altre parole, esistono «proprietà», intese nel senso estensionale, non esprimibili per mezzo di nessun dato linguaggio.

È proprio questa divergenza estensiva tra la nozione di «proprietà estensionale» e quella di «proprietà espressa nel dato linguaggio» che sta alla base della impossibilità di ritenere che i modelli delle teorie matematiche «ingenue», in particolare della teoria degli insiemi, riflettano tutti gli aspetti delle teorie ingenue.

Bisogna ancora notare che l'assioma del sottoinsieme è piuttosto uno «sche-

ma di assiomi». Esso definisce la serie infinita dei suoi casi particolari che si ottengono sostituendo alla proprietà indefinita P, una dopo l'altra, le frasi che esprimono tutte le possibili proprietà.

Per esempio, se, quale proprietà P dell'insieme u, prendiamo «u è un elemento di v», l'assioma del sottoinsieme postula l'esistenza della intersezione $y = x \cap v$.

Se, quale proprietà P, si prende la proprietà «u non è un elemento di v», noi postuliamo l'esistenza del *complemento* $y = x \setminus v$.

Introduciamo ancora alcune importanti nozioni utilizzando l'assioma del sottoinsieme ed altri assiomi.

Il prodotto cartesiano $x \times y$ degli insiemi x e y può essere definito cosí:

$$x \times y = \{z | z \in \mathcal{P}(\mathcal{P}(x \cup y)) \text{ e esistono } u \in x \text{ e } v \in y \text{ tali che } z = \langle u, v \rangle\}.$$

Pertanto $x \times y$ è costituito da tutte le coppie ordinate, il cui primo elemento appartiene ad x e il secondo a y. Osserviamo che, se $u \in x$ e $v \in y$, allora

$$\{u, v\} \in \mathcal{P}(x \cup y) \text{ e } \langle u, v \rangle = \{\{u\}, \{u, v\}\} \in \mathcal{P}(\mathcal{P}(x \cup y)).$$

Si chiama relazione binaria fra gli insiemi x, y ogni sottoinsieme di $x \times y$.

Si chiama applicazione f dell'insieme x in y (notazione: $f : x \to y$), oppure funzione dall'insieme x all'insieme y una relazione binaria $f \subseteq x \times y$ tale che per ogni $u \in x$ esiste un unico elemento $v \in y$ con la proprietà $\langle u, v \rangle \in f$. In tal caso spesso si scrive anche $v = f(u)$.

ZF8. ASSIOMA DELLA SOSTITUZIONE. *Sia P una proprietà di coppie ordinate di insiemi tale che per tutti gli elementi x, y, z di un certo insieme u, se le coppie $\langle x, y \rangle$ e $\langle x, z \rangle$ godono della proprietà P, risulti $y = z$. Esiste allora un insieme v tale che $y \in v$ se, e solo se, esiste $x \in u$ tale che la coppia $\langle x, y \rangle$ gode della proprietà P.*

I presupposti relativi a P sono piú deboli dei presupposti che definiscono una funzione. Ciò nonostante l'assioma postula sostanzialmente il fatto che P definisce una funzione su un conveniente insieme. Osserviamo che anche questo è uno schema di assiomi.

ZF9. ASSIOMA DELLA RESTRIZIONE. *Sia P una certa proprietà degli insiemi e supponiamo che esista un insieme con la proprietà P. Esiste allora un insieme x con la proprietà P tale che nessuno dei suoi elementi gode della proprietà P.*

È proprio da questo assioma che si può ricavare l'affermazione secondo la quale ogni catena $x_1 \ni x_2 \ni \ldots$ deve essere finita.

ZF10. ASSIOMA DELLA SCELTA. *Per ogni insieme x non vuoto esiste una funzione*

$$f : \{y | y \in \mathcal{P}(x) \text{ e } y \neq \emptyset\} \to x$$

tale che per ogni $y \in \mathcal{P}(x)$, $y \neq \emptyset$, si ha: $f(y) \in y$.

In altre parole si postula l'esistenza di una funzione che sceglie un solo elemento da ogni sottoinsieme non vuoto dell'insieme x.

Gli assiomi ZF1-ZF10 formano il sistema completo di assiomi della teoria degli insiemi secondo Zermelo-Fraenkel nel linguaggio naturale. Esistono altre varianti di assiomatizzazione della teoria degli insiemi. Fra queste la piú nota è l'assiomatizzazione di Gödel-Bernays. In essa la nozione fondamentale è la «classe», mentre l'insieme è definito quale classe che può essere elemento di un'altra classe. Nella interpretazione standard nell'universo di Zermelo-Fraenkel tutti gli insiemi formano una classe, ma non un insieme. Ogni proprietà degli insiemi, esprimibile nel linguaggio, definisce la classe di tutti gli insiemi che godono di questa proprietà. Questa classe pure non può essere un insieme. L'assiomatica di Gödel-Bernays non contiene schemi di assiomi e, pertanto, è piú comoda dal punto di vista formale. Per il contenuto essa può essere utile nella moderna teoria delle categorie; per esempio, la categoria di tutti i gruppi forma una classe, ma non un insieme. È noto, tuttavia, che nella teoria di Gödel-Bernays sono deducibili esattamente gli stessi risultati sugli insiemi della teoria di Zermelo-Fraenkel, cosí che esse in sostanza sono equivalenti.

Esaminiamo ora le nozioni metamatematiche introdotte nel paragrafo precedente applicate al linguaggio e agli assiomi della teoria degli insiemi.

5. *Caratteristica metateorica della teoria degli insiemi.*

La semantica del linguaggio della teoria degli insiemi nella variante iniziale di Georg Cantor [1895-97] era legata alla rappresentazione mentale di oggetti di natura arbitraria del mondo reale, o del mondo delle idee, e alla possibilità di immaginare collezioni diverse di tali oggetti come un tutto unico. Ognuna di queste collezioni è un insieme (termine standard scelto tra diverse espressioni sinonime del linguaggio naturale) e gli oggetti che ne fanno parte sono gli elementi di questo insieme. Da un tale punto di vista si potrebbe considerare, per esempio, l'insieme di tutti gli insiemi. Naturalmente esso deve contenere se stesso quale suo elemento.

A Bertrand Russell appartiene il ragionamento che mostra come si possa giungere assai presto ad una contraddizione applicando a questo arco di rappresentazioni mentali semplici costruzioni logiche. Infatti, consideriamo l'insieme X i cui elementi siano, tutti e soli, gli insiemi che non contengono se stessi quali propri elementi; per brevità li chiameremo insiemi «regolari». Poniamoci la domanda: è «regolare» X? Se X fosse «regolare», esso, per la definizione di «regolarità», non dovrebbe contenere se stesso quale suo elemento, ma ciò contraddirebbe la definizione di X quale insieme di *tutti* gli insiemi «regolari». D'altro lato, se X fosse «non-regolare», esso, per la definizione di «non-regolarità», dovrebbe contenere se stesso quale suo elemento, ma ciò contraddirebbe la definizione di X quale insieme che contiene *solo* insiemi «regolari».

Questo «paradosso» e quelli ad esso imparentati furono oggetto di numerose discussioni e disaccordi. Il modo meno radicale di uscirne fuori consistette nel restringere la semantica iniziale: bisognava ammettere che fossero insiemi non tutte le immaginabili collezioni di oggetti, ma soltanto alcune. Quello piú

radicale, invece, consistette nella richiesta di rinunciare completamente alla semantica del linguaggio della teoria degli insiemi e, come corollario, allo stesso linguaggio. Poiché tale richiesta distrugge completamente l'edificio della matematica classica e della maggior parte di quella contemporanea, pochi l'accolsero.

La semantica del linguaggio di Zermelo-Fraenkel, descritta all'inizio del paragrafo precedente, è il risultato di quella restrizione sulla rappresentazione degli insiemi che ci salva dal paradosso di Russell e, nello stesso tempo, permette di realizzare tutte le costruzioni matematiche. La collezione di Russell degli insiemi X che non contengono se stessi quali propri elementi, nell'universo di Zermelo-Fraenkel, è costituita da tutti gli insiemi dell'universo. Essa non è un insieme dell'universo; altrimenti avremmo $X \in X$, da cui si otterrebbe immediatamente la catena infinita di inclusioni $X \ni X \ni X \ni \ldots$, che in questo universo è impossibile.

Ciò che è stato detto non significa che questa semantica standard sia completamente accettabile.

Da un punto di vista puramente logico non è dimostrato che nel sistema ZF non si possa giungere ad una contraddizione mediante ragionamenti forse notevolmente piú lunghi di quello di Russell. Hilbert congetturò che la contraddittorietà degli assiomi potesse essere dimostrata con metodi puramente finiti, considerando la teoria combinatoria dei testi nella versione formalizzata del linguaggio della teoria degli insiemi. Ma, come dimostrò Gödel, in ognuna di tali dimostrazioni bisogna usare mezzi molto potenti: la formalizzazione di questa dimostrazione, quale deduzione dallo stesso sistema di assiomi, è impossibile. Ciò, per quanto concerne la non-contraddittorietà del sistema di assiomi di ZF, conduce a un circolo vizioso.

Da un punto di vista piú pragmatico è chiaro che l'universo di ZF contiene insiemi infiniti immensamente grandi, un'idea dei quali probabilmente è oltremodo lontana dalla realtà fisica. Secondo la critica di Brouwer non vi è alcun fondamento di aspettarci che sia giustificato, o almeno sia sensato, applicare a questi insiemi la logica standard astratta dalle considerazioni di collezioni finite. Ciò conduce a forti dubbi sul fatto che a testi arbitrari nel linguaggio di ZF sia applicabile l'idea della loro *verità-falsità* persino indipendentemente dal fatto che si sia o meno in grado di stabilire questa verità, o falsità.

Quest'ultima circostanza fu resa palese in modo particolarmente clamoroso con le dimostrazioni di Gödel e di Cohen, secondo le quali nel sistema di ZF non è deducibile dagli assiomi né l'ipotesi del continuo di Cantor, né la sua negazione. Se si crede che essa sia vera, o falsa, è assolutamente incomprensibile su quali principî ciò possa essere stabilito e in che misura risultino accettabili tali principî. In ogni caso gli assiomi di ZF non possono costituire un sistema completo.

Infine gli assiomi di ZF sono notoriamente non categorici. Infatti, Skolen ha dimostrato che ogni teoria formale, che ammetta generalmente una interpretazione, può essere interpretata nell'universo costituito semplicemente dai numeri interi e, per di piú, in infiniti modi sostanzialmente diversi tra loro.

In queste condizioni la maggioranza dei matematici assume la teoria degli insiemi, il suo linguaggio e la sua assiomatica, quale apparato per comunicare, fissare, scoprire e applicare i fatti matematici. Inoltre, con l'assunzione di questo linguaggio, si è aperta la possibilità di presentare la matematica come un tutto unico e di indagare tra le sue parti eterogenee una enorme quantità di legami stretti che precedentemente sfuggivano all'attenzione. Da questo punto di vista nessuna delle alternative proposte può competere con la teoria degli insiemi.

Dobbiamo limitarci a constatare che lo sviluppo della matematica non ha rivelato contraddizioni nella versione della teoria degli insiemi che è stata assunta; inoltre che l'esperienza storica ci testimonia la possibilità, quando se ne ravvisi la necessità, di precisare e correggere le nostre idee sugli oggetti matematici senza, per questo, demolire alla radice l'edificio matematico.

6. *Limiti del metodo assiomatico.*

Abbiamo già notato all'inizio dell'articolo il posto e il ruolo del metodo assiomatico sul piano teoretico generale. Le sue realizzazioni concrete nei diversi dominî del sapere formano tutto uno spettro, ad un estremo del quale vi sono le generalizzazioni teoretiche delle scienze umanistiche (probabilmente la tradizione è cominciata con l'*Ethica* dello Spinoza), mentre all'altro estremo vi sono le teorie formali della matematica moderna.

Tuttavia, accanto allo straordinario sviluppo delle idee e della tecnica di questo metodo, nel XX secolo fu scoperta l'esistenza di limiti di principio sulle sue possibilità che stanno al fondo della questione.

In questo campo la scoperta centrale è stata il *teorema di Gödel sulla incompletezza degli assiomi dell'aritmetica* [1931], il risultato e la dimostrazione del quale ammettono generalizzazioni significative. Descriviamo brevemente il contenuto di questo teorema.

Consideriamo un linguaggio L sufficientemente formalizzato dell'aritmetica, che comprenda i mezzi per esprimere le operazioni elementari sui numeri interi, le nozioni logiche fondamentali e i segni per i numeri interi concreti e indeterminati.

Numeriamo tutti i testi di questo linguaggio regolarmente costruiti con numeri naturali, per esempio, secondo il principio di Leibniz-Gödel. L'unica condizione essenziale in tale «numerazione di Gödel» consiste nel fatto che la ricostruzione del testo a partire dal numero e del numero a partire dal testo deve essere una procedura puramente meccanica, la cui attuazione possa essere affidata ad un calcolatore, avendo scritto per esso il programma corrispondente.

Fissiamo anche un sistema di assiomi dell'aritmetica. Si suppone che dal numero del testo si possa riconoscere meccanicamente se esso è, oppure no, un assioma.

Infine fissiamo nel linguaggio le regole d'inferenza. Anche la procedura per costruire, a partire dagli assiomi, tutti i testi deducibili deve essere meccanicizzata.

Si può dimostrare che in un'adeguata formalizzazione dei reali ragionamenti matematici tutte queste ipotesi sono soddisfatte. Di piú, si può argomentare il punto di vista consistente nel fatto che queste ipotesi devono essere soddisfatte in *ogni* adeguata formalizzazione.

Indichiamo ora con T l'insieme dei numeri dei testi veri e con D l'insieme dei testi deducibili dagli assiomi. È naturale ritenere che $D \subseteq T$. Il teorema di Gödel afferma che *questa inclusione è propria*, cioè *esiste sempre una affermazione vera dell'aritmetica non deducibile dagli assiomi.*

La sua dimostrazione è basata sull'unione di due principî.

ESPRIMIBILITÀ DI D: *esiste un testo P su L, comprendente il nome k di una variabile numerica intera, tale che $k_0 \in D$ se, e solo se, questo testo diventa vero dopo aver sostituito in esso il nome k con il numero intero k_0.* (Diciamo che questo testo P *esprime* l'insieme D).

NON ESPRIMIBILITÀ DI T: *per T un tale testo non esiste.*

La dimostrazione del primo principio richiede un lungo lavoro tecnico, ma sostanzialmente sfrutta semplicemente la possibilità della generazione meccanica degli elementi di D uno dopo l'altro. Questa generazione può essere affidata a un calcolatore e la sua logica e la sua aritmetica si prestano a una completa descrizione nel linguaggio L dell'aritmetica elementare.

La dimostrazione del secondo principio è un impiego molto ingegnoso dell'antico paradosso del bugiardo. Essa è fondata sul fatto che, per ogni testo P, si può costruire esplicitamente un nuovo testo Q_P, il cui significato è riferito con la descrizione seguente: *Q_P afferma che il numero di Q_P non appartiene all'insieme espresso dal testo P.*

Usiamo ciò per stabilire la non-esprimibilità di T. Supponiamo che T sia espresso da un testo P e giungeremo a una contraddizione. Per questo poniamoci la questione della veridicità del testo Q_P.

Se il testo Q_P è vero, dal suo significato segue che il numero di Q_P non appartiene all'insieme espresso dal testo P, cioè all'insieme T. Ma allora, per la definizione di T, il testo Q_P deve essere falso.

Se invece il testo Q_P è falso, dal suo significato segue che il numero di Q_P appartiene all'insieme espresso dal testo P, cioè all'insieme T. Ma allora, per definizione di T, il testo Q_P deve essere vero.

Ciò completa la dimostrazione del teorema di Gödel.

La non-esprimibilità di T fu dimostrata da Tarski. Gödel usò un metodo piú particolare che conduce a un risultato piú preciso. Egli, quale testo P, prese il testo che esprime l'insieme D. In tal caso il testo Q_P afferma che «il numero di Q_P non appartiene a D», cioè «Q_P non è deducibile dagli assiomi». Il testo Q_P non può essere falso, altrimenti risulterebbe che Q_P è deducibile dagli assiomi e perciò vero. Ne segue che il testo Q_P è vero e, per il suo significato, non è deducibile dagli assiomi. Pertanto il metodo di Gödel dà luogo a un infinito completamento degli assiomi dell'aritmetica con testi del tipo Q_P senza che, con ciò, ad ogni tappa di questo processo il sistema di assiomi risulti completo.

Per abbracciare tutta la verità solo sui numeri interi sono necessarie infinite

nuove idee. Per quanto ricco sia il sistema di premesse da cui si parte, questa verità non può essere generata meccanicamente. In misura maggiore queste limitazioni sono applicabili in altri dominî del sapere.

I successi della matematica e di settori matematicizzati delle scienze naturali indussero molti profondi pensatori a sperare nell'esistenza di alcune leggi universali, dalle quali tutte le altre verità potessero essere semplicemente dedotte teoricamente. Nella tradizione europea queste speranze sono legate ai nomi di Leibniz e Cartesio e sono tuttora sostenute da parte di alcuni fisici che meditano sulla struttura della nostra conoscenza della natura.

Tuttavia, dopo i lavori di Gödel, possiamo essere certi che queste speranze sono infinite. Anche lasciando da parte la questione di quanto sia complesso il mondo, sappiamo che il metodo delle inferenze deduttive non è sufficientemente forte.

Benché finora non siano stati esplicitamente indicati «problemi insolubili» (esempi di Gödel, ipotesi del continuo, problema dell'identità delle parole nei gruppi), lontani da quelli matematici, che stiano di fronte ai fisici, non è escluso che tali problemi fisici siano già stati affrontati. È possibile, per esempio, che le difficoltà di dedurre dalle equazioni della idrodinamica le caratteristiche di certi fenomeni realmente osservate siano effettivamente insuperabili: gli assiomi della teoria degli insiemi e l'apparato standard per tale deduzione possono non bastare.

In queste condizioni diventano particolarmente chiari il ruolo dell'esperimento numerico e dell'osservazione e il ruolo dell'intuizione e della congettura con insufficiente fondamento. Essi sono chiamati non solo a compensare la nostra provvisoria incapacità di raggiungere la verità mediante inferenze deduttive, ma forse anche a farci sapere che è impossibile conoscere la verità deduttivamente. Prima o poi dovremmo uscire dai limiti della «ragione pura».

Questi limiti di principio del metodo assiomatico, connessi con l'incompletezza dell'informazione da esso fornita e con il dubbio permanente sulla sua verità e non-contraddittorietà, fanno parte del fondamento della moderna gnoseologia e devono essere esplicitamente riconosciuti. Solo in queste condizioni può manifestarsi tutta la forza di quell'unica arma dell'intelletto umano che è il metodo assiomatico. [JU. I. M.].

Bourbaki, N.
1939sgg. *Eléments de mathématique*, Hermann, Paris.

Cantor, G.
1895-97 *Beiträge zur Begründung der transfiniten Mengenlehre*, in *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*, Springer, Berlin 1932.

Fraenkel, A. A.
1928 *Einleitung in die Mengenlehre*, Springer, Berlin 1928³.

Gödel, K.
1931 *Über formal unentscheidbare Sätze der Principia mathematica und verwandter Systeme*, in «Monatshefte für Mathematik and Physik», XXXVIII, pp. 173-98 (trad. it. in E. Agazzi, *Introduzione ai problemi dell'assiomatica*, Milano 1961).

Hilbert, D.
1899 *Grundlagen der Geometrie*, Teubner, Leipzig-Berlin 1930[7].

Streater, R. F., e Wightman, A. S.
1964 *PCT, Spin and Statistics and All That*, Benjamin, New York.

Zermelo, E.
1908 *Untersuchungen über die Grundlagen der Mengenlehre*, in «Mathematische Annalen», LXV, pp. 261-81.

Assiomi e postulati sono gli elementi fondamentali a partire dai quali, all'interno di teorie matematiche (cfr. **matematiche**), o piú in generale di teorie ad esse assimilabili, si procede deduttivamente per ottenere i risultati ricercati (cfr. **strutture matematiche, deduzione/prova, geometria/topologia** ed anche **fisica, quanta**). A fondamento della matematica è posta abitualmente la teoria degli insiemi (cfr. **insieme**); per evitare antinomie e paradossi è opportuno vagliare accuratamente gli assiomi iniziali ed agire all'interno di **un linguaggio** formalizzato (cfr. **formalizzazione**). Nell'articolo che precede si presenta come esempio la teoria di Zermelo-Fraenkel (cfr. **applicazioni**). Attraverso il teorema di Gödel, si accenna ai limiti intrinseci alle teorie formalizzate (cfr. **logica, referenza/verità**). Il rapporto tra teorie e teorie formalizzate è esaminato in particolare in **teoria**; piú in generale cfr. anche **scienza**. La «difficoltà» dell'accordo tra teorie formalizzate da una parte e dialettica dall'altra è esaminata in **dialettica**.

·io alla sussi-
·destinata a
· interno
·lla ri-
·elle
·e

Continuo/discreto

La categoria continuo/discreto ha una caratterizzazione gnoseologica molto generale. Il discreto è legato all'isolamento, alla separabilità degli oggetti e dei fenomeni, alla possibilità di fissarli per mezzo di simboli in una comprensione individuale e collettiva. Il modello matematico fondamentale del discreto è la successione dei numeri naturali, 1, 2, 3, ... Il continuo è legato alla possibilità di variazioni arbitrariamente piccole di questa o quella caratteristica, all'assenza di frontiere nette, di salti, alla connessione, ecc. Il modello matematico fondamentale del continuo è la retta reale, cioè il continuo dei numeri reali.

Tra le piú antiche e note considerazioni sul continuo/discreto vi sono i paradossi di Zenone di Elea (v secolo a. C.). Essi ci sono giunti nell'esposizione di Aristotele. Due, la «freccia» e «Achille», argomentano probabilmente l'impossibilità del moto in uno spazio-tempo continuo; prima di arrivare alla fine di un percorso un oggetto deve raggiungere la metà di questo percorso; quindi la metà del percorso rimanente e cosí via fino all'infinito. Altri due paradossi argomentano l'impossibilità del moto in uno spazio discreto: invero un oggetto in moto dovrebbe trovarsi e non trovarsi in uno stesso luogo.

Nonostante tutta l'influenza emozionale che i paradossi di Zenone hanno esercitato sui filosofi fino ai nostri tempi, bisogna subito convenire che non esistono né una loro interpretazione univoca, né un'unica idea su come li si dovrebbe risolvere. Il tentativo di trapiantare singoli frammenti dell'antica cultura umanistica su basi moderne è quasi sempre accompagnato da tali fenomeni d'incompatibilità. Perciò, per gli scopi di questo articolo, convenzionalmente, possiamo ritenere che nei paradossi di Zenone furono posti alcuni problemi, la cui soluzione era compito delle generazioni future [cfr. anche Grünbaum 1967].

Nel primo di essi si afferma l'*inconcepibilità* del movimento nell'ambiente spazio-temporale, sia continuo sia discreto. Benché nel testo di Zenone-Aristotele non si parli esplicitamente di ciò, si può supporre che sia proprio questo il senso dell'affermazione sull'*impossibilità* del moto. Zenone fu allievo di Parmenide, che divulgò, probabilmente, l'identità tra il pensiero e l'essere, benché l'interpretazione della sua filosofia sia altrettanto non-univoca.

Se si accetta quest'interpretazione di tali paradossi, si può dire che, per i suoi tempi, egli aveva ragione. Il movimento era inconcepibile nel senso che non esistevano né un apparato concettuale, né un linguaggio del pensiero sufficientemente sviluppati. Egli perciò rese evidente l'inadeguatezza e la povertà del linguaggio corrente in ordine a questo problema. L'apparato che occorreva fu elaborato in modo coerente da Aristotele, da Eudosso e da Euclide. Galileo e Newton iniziarono poi la nuova era moderna. Il moto meccanico negato da Zenone e forme di moto incomparabilmente piú complesse, che egli nemmeno immaginava, sono descrivibili nel linguaggio della matematica e della fisica, che tiene conto tanto degli aspetti continui quanto di quelli discreti. Il moto divenne definitivamente concepibile.

Tuttavia il linguaggio della matematica e della fisica moderne ha una logica interna straordinariamente complessa e, in quegli aspetti ove ci si riferisce alle astrazioni dell'infinito e del continuo, la sua semantica è tutt'altro che chiara. L'infinito matematico e la continuità, che pure sono dotati di un'enorme forza chiarificatrice nelle scienze naturali, in nessun modo possono essere interpretati direttamente in termini dei dati sperimentali di queste scienze. Perciò si può ritenere che le difficoltà presenti nei fondamenti stessi della matematica sono le lontane eredi dei paradossi di Zenone. È, questo, il secondo problema che egli pose. I suoi sono un modello dei moderni paradossi dell'infinito e, probabilmente, gli argomenti degl'intuizionisti e dei costruttivisti potrebbero essere vicini, nello spirito, a quelli di Zenone.

In quest'articolo ci si limiterà a descrivere il primo aspetto, sforzandosi di seguire come si presenta il continuo/discreto nei dati sperimentali delle moderne scienze naturali e nei loro modelli teorici.

Il primo paragrafo è dedicato alla fisica e ai suoi modelli; il secondo, ai processi d'informazione e al linguaggio nel senso piú esteso della parola; il terzo, ad alcuni modelli matematici complementari del continuo/discreto, l'importanza dei quali nelle scienze naturali non è ancora chiara, ma non è improbabile che aumenti.

1. *Spazio, tempo, materia, energia.*

In questo paragrafo si vedrà in che modo si manifestano gli aspetti del continuo/discreto nel quadro generale del mondo cosí come esso è delineato dai dati della fisica moderna. Questo quadro non è omogeneo; si presenta come una gerarchia di numerosi livelli, ognuno definito dalle sue peculiari unità di misura spazio-temporali. Per descrivere la natura a questi livelli sono stati elaborati diversi modelli fisico-matematici, nei quali a volte si sottolineano le caratteristiche continue del quadro osservato, altre volte quelle discrete; oppure esse vi vengono sintetizzate, come accade nell'apparato della meccanica quantistica. La frammentarietà delle attuali conoscenze fisiche è una circostanza ben nota: non abbiamo una teoria che possa pretendere, anche solo provvisoriamente, di descrivere universalmente e conseguentemente la natura a tutti i suoi livelli, benché si creda che i principî della meccanica quantistica siano applicabili universalmente. Uno dei problemi chiave nei tentativi di descrizione unitaria del macro- e del micromondo consiste appunto nella difficoltà di far concordare il quadro geometrizzato dello spazio-tempo macroscopico, in cui domina l'aspetto della continuità, con il quadro quantistico del micromondo, dove hanno un ruolo fondamentale il discreto e la discontinuità dei microfenomeni e il carattere probabilistico dei salti quantici.

Piú oltre si esporranno brevemente, nei limiti del necessario, la fenomenologia e i modelli dei diversi livelli del mondo osservato.

1.1. Fenomenologia.

« A misura d'uomo ». Le unità di misura standard di lunghezza, tempo e forza, cioè centimetro, secondo e grammo, corrispondono all'ordine di grandezza del corpo umano e alla qualità della nostra esperienza quotidiana. Tra gli ordini di grandezza 10^{-2} e 10^{4} cm si colloca la maggior parte degli oggetti naturali e artificiali con i quali l'uomo ha avuto a che fare nel corso della maggior parte del suo sviluppo: oggetti d'uso quotidiano, strumenti di lavoro, abitazioni, insetti, fauna e flora. Le prime idee sul discreto si formarono con la consapevolezza che questi oggetti sono separati l'uno dall'altro, dotati di limiti spaziali e di esistenze temporali distinte. Queste idee si sono concretizzate in quelle conquiste della cultura umana quali il «far di conto» con i numeri naturali (all'inizio piccoli), il linguaggio articolato e i rudimenti della scrittura; nei sistemi di comunicazione dell'uomo apparvero i segni sonori, grafici ed altri (cfr. oltre).

L'intorno spaziale piú vicino all'uomo, quello accessibile direttamente alla conoscenza mediante gli organi sensoriali (eccetto la vista, le cui possibilità sono incomparabilmente maggiori), possiede le stesse dimensioni caratteristiche. La parte «vuota» (cioè riempita dall'atmosfera) di questo intorno, dove è possibile, senza impedimenti, il moto dell'uomo, degli animali e degli oggetti – il moto che si sviluppa nel tempo e il cui corso non presenta salti osservabili e frontiere – tutto ciò fu alla base delle prime idee sulla continuità. Occorsero decine di millenni di sviluppo dell'umanità affinché queste idee si formulassero nelle nozioni di «spazio vuoto» assolutamente continuo, omogeneo, uniforme, e di «tempo vuoto» che dall'epoca di Newton restano una delle fondamentali generalizzazioni della fisica.

Lo spazio reale sopra la superficie della Terra è interrotto da villaggi, fiumi, boschi e monti. Il tempo reale è contrassegnato dal succedersi di notti, giorni, stagioni, generazioni. Questi processi naturali furono all'origine del calcolo e delle misurazioni, cioè delle primitive sintesi del discreto e del continuo.

In confronto alle dimensioni naturali del corpo umano, nello spazio e nel tempo, le possibilità energetiche dei suoi organi sensoriali presentano un aspetto sostanzialmente nuovo. Quella che è in pratica la piú piccola unità di lavoro derivata dalle unità fondamentali, cioè il grammo-peso, si trova effettivamente quasi al limite inferiore delle facoltà meccaniche dell'uomo. Invece i suoi organi sensoriali, e tra essi l'udito e in particolare la vista, sono in grado di percepire e distinguere quantità di energia incomparabilmente piú piccole. Secondo indagini moderne l'occhio, completamente adattato al buio, reagisce in modo sicuro a una quantità di luce corrispondente a cinque o sei fotoni, che sono le piú piccole unità discrete di energia luminosa. Solo per questo siamo in grado di vedere oggetti cosí lontani come le stelle. E, già da molto tempo, l'elaborazione delle rivelazioni dell'occhio da parte del cervello ci ha permesso di sentirci parte di un mondo immensamente piú grande di noi. Tuttavia la comprensione che l'energia luminosa e quella acustica hanno natura comune con quella meccanica è un'acquisizione molto tarda della cultura umana: le prime dimostrazioni sicure di questo

fatto sono state ottenute venti-trent'anni fa. Il ruolo dei nostri organi di senso va ben oltre quello di «misuratori», o «trasduttori», di lavoro; essi ci procurano un'enorme quantità d'informazioni che, per valore, è incomparabile con le misure energetiche dei deboli flussi di azioni. Agli aspetti continui/discreti dei processi d'informazione è dedicato in modo speciale il paragrafo seguente.

Dimensioni astronomiche e cosmologiche. L'osservazione dei corpi celesti risale alle primissime tappe della civilizzazione umana. Tuttavia la comprensione della loro natura e delle vere dimensioni spaziali e temporali della loro esistenza è frutto degli ultimi secoli.

La Terra è approssimativamente una sfera di raggio 10^9 cm. Il corpo celeste piú importante per i processi terrestri è il Sole, che si trova a una distanza di $1{,}5 \cdot 10^{13}$ cm. Il sistema solare nel complesso ha un diametro cento volte piú grande: da Plutone ci divide una distanza di circa $2 \cdot 10^{15}$ cm. La stella piú vicina è a una distanza di circa $4{,}5 \cdot 10^{18}$ cm. Una comoda unità di lunghezza per misurare le distanze infrastellari è l'anno luce, che è lo spazio percorso dalla luce in un anno. Un anno luce è circa 10^{18} cm.

Il Sole, il sistema planetario e le stelle ad esso piú vicine sono una parte insignificante di un enorme collettivo di stelle e nebulose che insieme formano la nostra Galassia, la cui massa fondamentale costituisce la Via Lattea. Grosso modo la Galassia è a forma di lente biconvessa con un diametro intorno ai 10^5 anni luce e uno spessore di circa 1500 anni luce. Essa contiene circa 10^{11} stelle. Noi ci troviamo alla periferia della Galassia, a circa $3 \cdot 10^4$ anni luce dal suo nucleo e a 30 anni luce dal suo piano equatoriale. Nella Galassia le stelle sono distribuite in modo assai difforme; la densità al centro è decine di migliaia di volte quella intorno a noi. Gli oggetti come la nostra Galassia si contano attualmente a miliardi e sono riuniti, pare, in un enorme sistema, la Metagalassia. Le galassie piú vicine formano le Nubi di Magellano e sono lontane da noi circa $2 \cdot 10^5$ anni luce, cioè una distanza il cui ordine di grandezza coincide con le dimensioni della nostra Galassia. Le piú lontane galassie osservate si trovano a una distanza di miliardi di anni luce. Nel 1929 Hubble determinò le distanze di una ventina di galassie e ne misurò le relative velocità radiali. Risultò che tutte queste galassie si allontanano da noi con una velocità approssimativamente proporzionale alla distanza. Il coefficiente di proporzionalità, cioè la costante H di Hubble, in base alle misurazioni attuali, ammonta a $55 \cdot 10^5$ cm/sec per ogni $3{,}3 \cdot 10^6$ anni luce di distanza. Tra l'altro questa cifra è stata piú volte rivista. Estrapolando questo moto a tutta la parte visibile dell'Universo e al lontano passato, si ricava che circa 10^{10} anni fa l'Universo era concentrato in uno spazio molto piccolo; poi cominciò ad espandersi in modo simile a un'esplosione e questa espansione è tuttora in atto.

Nelle dimensioni cosmiche e in prima approssimazione, il nostro mondo si presenta pertanto, nel continuo spazio-temporale, come evoluzione di materia, distribuita in modo niente affatto uniforme, che presenta caratteristiche di discretezza a diversi livelli dimensionali. I corpi celesti sono oggetti nettamente distinguibili; in essi la densità della materia oscilla fondamentalmente tra i limiti di 10^{-10}-10^3 g/cm^3, mentre la distanza fra le stelle, per esempio nell'intorno del

Sole, supera 10^7 volte il loro raggio. A sua volta, nelle dimensioni delle galassie (10^{25} cm), la densità media della materia è dell'ordine di 10^{-23} g/cm^3, mentre a livello delle distanze fra le galassie essa è inferiore di tre-quattro ordini. Infine, nelle dimensioni di 10^{26}-10^{27} cm, la distribuzione media della materia diventa come se fosse omogenea e isotropa. Lo spazio cosmico che divide i coaguli di materia non è niente affatto vuoto; oltre alla polvere interstellare e ai gas, in esso vi sono i campi. Le oscillazioni del campo elettromagnetico, cioè la luce e le radioonde, sono per noi la sorgente fondamentale delle conoscenze sul mondo. I campi gravitazionali, cioè i campi delle forze di gravità, legano tra loro i coaguli di materia e, secondo le concezioni della teoria generale della relatività di Einstein, definiscono la geometria generale dello spazio-tempo. Lo spazio è poi attraversato da flussi di radiazione cosmica e di neutrini. Essi sono portatori contemporaneamente di massa e di energia, legate tra loro dalla relazione fondamentale di Einstein $E=mc^2$ (in cui E è l'energia, m la massa e c la velocità della luce). La descrizione macroscopica dei campi sottolinea il loro aspetto continuo. Cosí, il campo elettromagnetico è descritto dalla posizione dei vettori delle intensità delle componenti elettrica e magnetica che variano nel tempo e nello spazio in accordo con le equazioni differenziali di Maxwell. La descrizione dei campi (e della materia) secondo la meccanica quantistica sottolinea l'aspetto discreto: il campo elettromagnetico è descritto quale «gas di fotoni», quanti di radiazione che si muovono con la velocità della luce. Su ciò si tornerà piú avanti.

Dimensioni atomico-molecolari. Nelle dimensioni di 10^{-3}-10^{-5} cm sono comprese le piú piccole unità della vita: le cellule, i microorganismi, i virus. La vita, quale oggetto delle scienze biologiche, è straordinariamente complessa e nel suo studio non ci siamo ancora allontanati da un livello puramente descrittivo. E, a un tale livello, in ogni scienza l'aspetto discreto è un portato della stessa necessità di classificare i fenomeni studiati, d'individuare le unità strutturali, gli organi, i processi elementari. Una delle piú grandi generalizzazioni della biologia fu l'individuazione delle cellule quali unità strutturali fondamentali di tutto ciò che vive. Dal punto di vista del tema di questo articolo i dati biologici piú interessanti si riferiscono, probabilmente, alla trasmissione delle informazioni nei sistemi viventi; ad essi sarà dedicato il paragrafo seguente. Qui, invece, tralasciando il livello biologico, conviene passare al micromondo della chimica e della fisica, ricordando anzitutto le idee fondamentali sugli atomi e le molecole.

Nell'aspetto macroscopico le «sostanze pure» dei chimici, come l'acqua, forniscono gli esempi piú tipici della continuità, della omogeneità e della divisibilità delle sostanze. La questione se questa divisibilità sia illimitata ha occupato da tempi remoti un posto importante nelle costruzioni della filosofia naturale. Il suo esame ha condotto già nell'antichità all'idea speculativa degli atomi quali piú piccole unità delle sostanze non ulteriormente divisibili. L'ipotesi sugli atomi forní una spiegazione qualitativa verosimile di effetti quali la compressibilità delle sostanze, l'esistenza delle soluzioni, la presenza dei tre stati solido, liquido e gassoso. Nella terminologia moderna si chiamano atomi le piú piccole unità di un elemento chimico, mentre le piú piccole quantità di sostanze pure, unione di ele-

menti, si chiamano molecole e sono costituite da atomi legati tra loro. Nel XIX secolo furono costruite le prime teorie quantitative, in cui le ipotesi sugli atomi e le molecole assunsero lo stato di concetti concordanti e dotati di una grande forza chiarificatrice. Fra esse vi è la «legge delle proporzioni multiple» per le sostanze pure: per esempio, in ogni volume d'acqua le componenti in peso comprendono due unità d'idrogeno e una unità (mole) di ossigeno. Ancora, l'ipotesi che un gas sia costituito da molecole che si muovono in modo caotico ha permesso a Maxwell e Boltzmann di chiarire teoricamente le leggi sperimentali sul comportamento dei gas. Probabilmente il culmine di questo periodo fu il calcolo del famoso numero di Avogadro, $6 \cdot 10^{23}$, che esprime la quantità di molecole presenti in una mole di qualunque gas. Questa costante cosí grande stabilisce il primo legame tra il micro- e il macromondo. Grazie alla sua grandezza è possibile uno dei piú diffusi meccanismi di formazione di macroeffetti continui quale risultato di una media su microeffetti discreti. È, questo, l'oggetto della fisica statistica.

Il dominio tra 10^{-6} e 10^{-7} cm si riferisce alle dimensioni molecolari: dalle piú grosse molecole proteiche all'atomo. Nelle indagini scientifiche reali si usa tutto uno spettro di modelli funzionali di molecole e di atomi. Ad un estremo di questo spettro vi sono veri e propri modelli (di legno, di plastica e di metallo) di grosse molecole, in cui gli atomi sono rappresentati da sferette rigide di dimensioni definite e sono rigidamente fissati nella molecola da vincoli infratomici, le cui lunghezze e angoli reciproci sono noti e costanti. Nei modelli piú realistici l'atomo è immaginato nella forma di un nucleo rigido del diametro di circa 10^{-13} cm circondato da una nube di elettroni, suddivisa in orbite elettroniche che possono deformarsi e riunirsi dentro la molecola, formando dei vincoli non rigidi a guisa di molle. Nelle molecole gli atomi possono oscillare attorno alla loro posizione media; i vincoli pure possono oscillare e ruotare. Modelli ancor piú realistici richiedono di passare a un linguaggio del tutto diverso: quello fortemente matematicizzato della meccanica quantistica che, a sua volta, ammette descrizioni con diversi gradi di precisione. In questo linguaggio una molecola, o un atomo, possono essere descritti da una funzione ψ a valori complessi (funzione d'onda o di stato), dipendente da una serie di parametri continui (le coordinate degli elettroni, dei nuclei e del tempo) e da una serie di argomenti discreti (per esempio, lo spin), la cui evoluzione tra un'interazione e un'altra è definita dall'equazione differenziale a derivate parziali di Schrödinger, e per effetto delle interazioni avvengono salti quantici probabilistici della funzione di stato ψ descritti da speciali regole supplementari (cfr. oltre).

Le proprietà degli atomi non risultarono quelle che ci s'immaginava nell'antichità e persino all'epoca di Newton.

Prima di tutto gli atomi sono divisibili; essi sono costituiti da un nucleo circondato da una nube di elettroni e, a sua volta, il nucleo è formato da protoni e neutroni. A livello atomico e subatomico entriamo in un ambito dominato nettamente dalle caratteristiche discrete della materia e dell'energia e ciò si nota particolarmente nelle descrizioni fenomenologiche. La loro sintesi con gli aspetti continui si raggiunge solo al livello fortemente matematicizzato della teoria quantistica. Ecco una lista di alcune fondamentali proprietà discrete.

Struttura del nucleo. Due caratteristiche numeriche fondamentali dell'atomo sono il numero di massa A, uguale alla quantità totale dei protoni (carichi positivamente) e dei neutroni (neutri) nel nucleo, e il numero atomico Z, uguale al numero dei protoni nel nucleo, che coincide con il numero degli elettroni nell'atomo. Le proprietà chimiche di un atomo sono essenzialmente definite dal numero atomico Z che può assumere i valori da $Z = 1$ (idrogeno) a $Z = 103$ (laurenzio) e, probabilmente, maggiori. Gli atomi con grandi valori di Z sono instabili. Per un dato Z il numero di massa A può assumere diversi valori in seguito alla variazione della quantità di neutroni nel nucleo; si hanno cosí i diversi isotopi di un elemento con un dato Z. Le proprietà chimiche degli atomi rivelano una determinata periodicità dipendente da Z, che fu scoperta prima di essere chiarita in base alla teoria atomica e fu alla base della disposizione degli elementi nella tabella di Mendeleev.

Nubi elettroniche. Un comodo modello della struttura della nube elettronica nell'atomo, che chiarisca in particolare la struttura discreta degli spettri atomici, parte dall'ipotesi che, al crescere di Z, nuovi elettroni riempiono gradatamente gli strati usualmente indicati con le lettere K, L, M, N, ... L'atomo dell'idrogeno contiene un elettrone nel K-strato; l'atomo dell'elio, due elettroni nel K-strato; l'atomo del litio, due elettroni nel K-strato e uno nell'L-strato. Il K-strato non può contenere piú di due elettroni; all'L-strato è diviso in due sotto strati, $2s$ e $2p$, il primo dei quali non può contenere piú di due e il secondo piú di sei elettroni. Nell'atomo del neon ($Z = 10$) il K-strato e all'L-strato sono pieni. Le proprietà chimiche di un elemento sono definite in misura notevole da quanto è riempito l'ultimo strato.

In una teoria piú raffinata la nube elettronica attorno al nucleo è descritta da una funzione di stato ψ dipendente da $3Z$ argomenti spaziali e da $2Z$ argomenti che assumono due valori (gli spin degli elettroni). Il quadrato del modulo della funzione di stato ψ fornisce la densità spaziale della nube elettronica e in questa densità non è cosí facile scorgere la struttura della nube. Infatti i numeri di elettroni che riempiono gli strati sono alcune caratteristiche matematiche di simmetria di questa funzione di stato ψ (si vedano alcuni particolari al § 1.3).

Livelli di energia. Ogni microsistema, quale l'atomo, la molecola, il nucleo, ecc., può trovarsi in stati energetici diversi caratterizzati in particolare dal livello di energia E «accumulata» nel sistema stesso. Lo stato fondamentale di un sistema corrisponde a quello con energia minima ed è il piú stabile. La descrizione delle nubi elettroniche data sopra si riferiva allo stato fondamentale dell'atomo. Sotto l'azione di un campo elettromagnetico un sistema può passare ad uno stato eccitato con energia piú grande, assorbendola dal campo, e, al contrario, può passare a un livello energetico inferiore, emettendo una parte di energia sotto forma di radiazione elettromagnetica.

I livelli di energia di tali sistemi, quali l'atomo e il nucleo, hanno componenti discrete e continue. Per esempio, i livelli di energia dell'atomo dell'idrogeno

sono $E_n = -R/n^2$, dove $n \geq 1$ è un numero intero e R è la costante di Rydberg. Lo stato fondamentale corrisponde a $n = 1$. Quando l'atomo passa dallo stato n allo stato m si ha un'emissione luminosa di energia $E_n - E_m$ (per $n > m$), oppure un assorbimento di energia $E_m - E_n$ (per $n < m$). Queste quantità di energia corrispondono alle linee spettrali. La posizione della linea è definita dalla frequenza della luce e il legame tra frequenza ed energia è dato dalla relazione $E = h\nu$, dove h è la costante di Planck. Le energie $E \geq 0$ formano una semiretta continua e corrispondono agli stati ionizzati dell'atomo dell'idrogeno, quando l'elettrone non è legato al nucleo.

La costante di Planck, $h = 6{,}626 \cdot 10^{-27}$ erg·sec, chiamata anche quanto d'azione, è una costante fondamentale che appare nella descrizione di tutte le proprietà discrete del micromondo. La maggior parte delle caratteristiche della materia e dell'energia, che nella teoria classica si immaginano distribuite in modo continuo e in quella quantistica in modo discreto, possono assumere valori proporzionali ad h. Poiché questa costante è molto piccola, in molte osservazioni macroscopiche tali livelli discreti risultano praticamente indistinguibili, ciò che fornisce una continuità approssimata. Da questo punto di vista il ruolo della costante di Planck è simile a quello del numero di Avogadro. Però, mentre quest'ultimo ammette un'interpretazione completamente classica in termini di «conteggio» delle singole molecole in un volume di gas, la costante di Planck, invece, riflette aspetti della struttura della materia essenzialmente nuovi e che, di fatto, non possono essere espressi nel linguaggio della fisica classica.

Consideriamo alcuni fenomeni, per spiegare i quali è necessario introdurre la costante di Planck.

Fotoni-quanti di radiazione elettromagnetica. La costante h è stata scoperta da Planck nel 1900 a conclusione dei tentativi di chiarire le leggi sperimentali della radiazione, cioè il carattere della dipendenza della lunghezza d'onda, con densità massima di radiazione, dalla temperatura del corpo irradiante. Risultò che la formula teorica esatta si ottiene se si suppone che l'atomo, considerato quale oscillatore con frequenza ν, può ricevere o emettere solo porzioni $h\nu$ di energia. Piú tardi tali porzioni furono chiamate fotoni (di frequenza ν). In ben noti limiti essi possono essere riguardati come particelle di massa nulla che si muovono con la velocità della luce. Un atomo allo stato eccitato può passare a quello fondamentale e in tal caso, passando un elettrone da una nube piú alta a una piú bassa, l'atomo emette un fotone. Al contrario, un atomo può assorbire un fotone e, come risultato, uno dei suoi elettroni passerà a un livello energetico superiore, oppure lascerà del tutto l'atomo. Quest'ultimo fenomeno si chiama effetto fotoelettrico (esterno) e la spiegazione delle sue leggi quantitative, dovuta ad Einstein (1905), fu la seconda fondamentale conferma dell'ipotesi di Planck.

I dati sperimentali suggerivano che, illuminando un metallo, l'energia degli elettroni espulsi dagli atomi dipende solo dalla frequenza («dal colore») della luce e non dalla sua intensità. In generale non si ha emissione elettronica finché la frequenza ν non supera un certo valore critico; dopo di che si ha l'emissione e l'energia degli elettroni che abbandonano gli atomi è della forma $E = h\nu - W$. La

causa di ciò sta nel fatto che, fino a quando i quanti della radiazione incidente sono minori dell'energia di fuga W, l'elettrone non può essere espulso, poiché esso assorbe energia solo in porzioni del valore di $h\nu$. Dopo di che, assorbendo un quanto, l'elettrone lascia l'atomo con un'energia uguale a $h\nu - W$. I quanti sono indivisibili.

Dimensioni nucleari e subnucleari. Essere penetrati nella struttura della materia nelle dimensioni al di sotto di 10^{-12} cm è il piú importante successo della fisica degli ultimi decenni. Ciò ha richiesto la creazione di nuovi mezzi tecnici e di un linguaggio della fisica nuovo e ancora ben lontano dall'essere completo. La stessa fenomenologia di questo vastissimo mondo, i cui dati sono ottenuti da osservazioni con l'aiuto di acceleratori ad alta energia, da intuizioni teoriche audaci, da sottili esperimenti e mediante lunghi calcoli automatici, è eccezionalmente complessa. Le nozioni fondamentali di questa fenomenologia sono le particelle elementari, le loro interazioni e i campi. Attualmente sono note alcune centinaia di particelle elementari e di esse e delle interazioni cui partecipano si ha una classificazione fenomenologica frammentaria.

Le particelle elementari sono attualmente le ultime unità strutturali note della materia-energia nel senso seguente.

La divisione della materia può essere spontanea, o risultato di un'interazione con passaggio di energia tra masse diverse di materia stessa. A livello macroscopico il modello di suddivisione si presenta come un frazionamento meccanico in seguito ad urti. A livello molecolare e atomico la divisione della materia avviene nei processi di vaporizzazione (energia calorifica), scomposizione chimica, ecc.

La separazione degli elettroni dagli atomi avviene, per esempio, nel riscaldamento dei metalli, oppure nei processi fotoelettrici (energia di radiazione elettromagnetica).

La divisione del nucleo può essere sollecitata da un suo bombardamento con particelle ad alta energia e può avvenire anche spontaneamente.

Ci si potrebbe attendere che ulteriori processi di divisione delle particelle elementari possano essere osservati, per esempio, come conseguenze di urti reciproci. Invece la situazione reale risulta qualitativamente diversa. Precisamente, se l'energia d'urto delle particelle non è grande, può semplicemente variare il carattere del loro moto («urto elastico»). Se invece è grande, cioè se, riferita alla massa secondo la formula $E = mc^2$, diventa confrontabile con la massa della particella elementare, al posto della sua rottura possono generarsi nuove particelle, quanti di radiazione, e cosí pure possono verificarsi varie mutazioni di particelle in altre. In modo molto approssimato si può dire che, quando l'energia d'urto necessaria per spezzare la particella è confrontabile con la massa della particella stessa, invece della rottura si verifica la creazione di nuove particelle, nelle quali si materializza questa energia d'urto.

Per esempio, l'urto di due protoni può produrre due protoni e due mesoni π (pioni), oppure i cosiddetti iperoni. Un quanto di radiazione può generare nel campo elettromagnetico di un atomo una coppia di particelle costituita da un elettrone e un positrone. Al contrario, questa coppia, in un urto, può annichilarsi

generando due (o piú) quanti di radiazione. Attualmente viene osservata sperimentalmente una quantità enorme di tali interazioni; sono studiate le probabilità di diversi «canali di reazione»; sono registrati i loro prodotti e tutto ciò costituisce la fenomenologia del mondo delle particelle elementari.

Le particelle elementari sono stabili in misura diversa. Protoni, elettroni e fotoni sono ritenuti assolutamente stabili. Un neutrone libero, invece, si scinde spontaneamente in un protone, un elettrone e un antineutrino in media dopo diciassette minuti, mentre un pione neutro vive circa 10^{-16} sec, dopo di che si scinde o in due fotoni (con una probabilità intorno a 0,988), oppure in un fotone, un elettrone e un positrone. A ogni particella elementare corrisponde un'antiparticella. È, questa, una manifestazione di una certa fondamentale simmetria delle leggi fisiche. L'elettrone e il positrone sono antiparticelle l'una dell'altra; l'«antifotone» coincide con il fotone; l'antiparticella del neutrone si chiama appunto antineutrone.

A seconda della massa a riposo le particelle sono suddivise nelle seguenti classi fondamentali:

a) fotoni, con massa a riposo nulla;
b) leptoni (particelle leggere): vi appartengono l'elettrone e il positrone, i mesoni μ, il neutrino e l'antineutrino;
c) mesoni: pioni, kaoni, mesoni η e altre;
d) barioni: protone e neutrone, gli iperoni e le loro antiparticelle.

Sono pure note moltissime particelle enormemente instabili, spesso trattate quali *risonanze*, cioè quali stati eccitati di particelle piú stabili.

Del comportamento delle particelle elementari negli urti, in sistemi tipo atomo e nucleo, e cosí pure dei decadimenti, sono responsabili tre tipi d'interazioni: elettromagnetiche, deboli e forti. Alle interazioni elettromagnetiche partecipano immancabilmente i fotoni, che sono i loro veicoli. Le interazioni deboli sono caratterizzate dal fatto che vi partecipa il neutrino, oppure dal fatto che non si conserva un particolare numero quantico, la «stranezza» (si veda piú avanti). Le interazioni forti sono caratteristiche dei processi che interessano il nucleo atomico e i loro veicoli sono pioni e altre particelle, che prendono il nome di adroni. Fotoni e leptoni non partecipano a interazioni forti.

Le interazioni tra particelle elementari sono soggette a leggi di conservazione di una serie di specifici numeri quantici discreti, che nei sistemi di particelle elementari sono additivi. Per esempio, in tutte le interazioni si conserva la carica elettrica totale, che assume valori multipli della carica dell'elettrone *e*. Inoltre, in tutte le interazioni si conserva il numero barionico, uguale a 0 per il fotone, i leptoni e i mesoni; a $+1$ per il protone, il neutrone e cinque iperoni; e a -1 per le corrispondenti antiparticelle. La conservazione del numero barionico spiega fenomenologicamente la stabilità del protone: essendo il piú leggero dei barioni, esso, in virtú di questa legge di conservazione, non può decadere in particelle piú leggere. Nelle interazioni forti ed elettromagnetiche si conserva anche la cosiddetta ipercarica, che però può non conservarsi nelle interazioni deboli. Infine si conserva sempre la carica leptonica.

Riassumendo, si può dire che nel mondo delle particelle elementari dominano le leggi di tipo discreto che definiscono lo spettro delle loro masse, i tipi d'interazione e di canali di reazione e le leggi di conservazione. A fondamento della meccanica quantistica vi sono le idee sulle funzioni d'onda (o di stato) e sui campi operatoriali che si evolvono nel continuo spazio-temporale e le caratteristiche discrete legate alle loro proprietà di simmetria, singolarità e risonanza che saranno discusse nel § 3.

1.2. Modelli teorici della macrofisica.

Nei modelli fondamentali della fisica del macromondo gli eventi hanno luogo nell'ambiente «continuo» dello spazio-tempo. I corpi materiali sono spesso sostituiti da particelle puntiformi, il cui moto è governato da equazioni differenziali ordinarie (meccanica). I mezzi continui sono modellati da campi scalari, vettoriali o piú complessi, la cui struttura spazio-temporale è definita da equazioni differenziali alle derivate parziali (teoria dei campi). I mezzi, quali i gas e i liquidi, possono essere riguardati come se fossero costituiti da un gran numero di particelle classiche (molecole) e sorge cosí il problema di ricavare delle equazioni «di campo» mediate a partire dalle equazioni meccaniche del moto delle singole particelle (fisica statistica). In questo ambiente continuo le caratteristiche discrete, o discontinue, del moto possono presentarsi sotto forma di non-omogeneità e di «stati legati» della materia, quali le stelle, il sistema solare, le transizioni di fase, oppure sotto forma di singolarità dei campi (valori che tendono all'infinito, discontinuità, onde d'urto), oppure sotto forma di risonanze, di pacchetti d'onda, o ancora sotto altre forme. In questo paragrafo si prenderanno brevemente in esame tali effetti.

Modelli dello spazio-tempo. In tutte le teorie fondamentali della fisica moderna lo spazio-tempo è modellato mediante varietà differenziabili a quattro dimensioni con questa o quella struttura complementare. Si ricordi (cfr. l'articolo «Applicazioni» di questa stessa *Enciclopedia*) che una tale varietà può essere definita assegnando: *a*) un insieme M di punti; *b*) un atlante su M costituito da una famiglia di carte, ognuna delle quali rappresenta un sistema di coordinate (x_1, x_2, x_3, x_4) che identifica un certo intorno di un punto di M con un certo insieme aperto di $\mathbf{R}^4$. Queste carte devono ricoprire tutto M e nell'intersezione di due carte il passaggio da un sistema di coordinate all'altro deve essere dato da funzioni differenziabili.

L'astrazione fisica corrispondente a un punto di M è un evento localizzato in un punto dello spazio e che avviene istantaneamente. L'urto fra due particelle elementari, o l'emissione di un quanto da un atomo, possono essere considerati approssimazioni di questa astrazione. La meccanica quantistica stabilisce alcuni limiti dell'applicabilità di quest'astrazione che si vedranno piú avanti; a dimensioni macroscopiche tali limitazioni possono ritenersi, fino a un certo grado, inessenziali.

L'astrazione fisica corrispondente a un sistema locale di coordinate in M è

una certa convenzione sui principî della misura delle grandezze spaziali e temporali in una regione dell'Universo, gli ordini delle quali possono variare tra limiti estesi da quello atomico e subatomico a quello di «tutto» l'Universo. È proprio la scelta di queste convenzioni, insieme con questi o quei principî d'invarianza delle leggi fisiche, che definisce le caratteristiche della geometria locale e globale dei diversi modelli dello spazio-tempo.

Spazio-tempo classici. Nelle dimensioni macroscopiche del laboratorio le distanze possono essere misurate con righe rigide e strumenti simili, mentre il tempo può essere misurato mediante processi periodici di questo o quel genere, in particolare con gli orologi. È naturale ritenere l'indicazione dell'orologio costante per tutti i luoghi del laboratorio, trascurando la velocità finita nella trasmissione dei segnali (della luce) del parte dell'orologio stesso, oppure calcolando una correzione dovuta al ritardo nella lettura dell'orologio. Ciò conduce a un modello dello spazio-tempo del laboratorio della forma $U \times T$, dove U è un dominio dello spazio euclideo e T è un intervallo della retta dei tempi. Estrapolando questo modello «a tutto il mondo e per l'eternità» giungiamo al modello $\mathbf{R}^3 \times \mathbf{R}$, che è dato da un'unica carta corrispondente a un dato sistema del laboratorio con le coordinate (x_1, x_2, x_3, t). Se $(x_1^{(i)}, x_2^{(i)}, x_3^{(i)}, t^{(i)})$, $i = 1, 2$, sono le coordinate di due punti dello spazio-tempo in questo sistema, risultano definiti gli intervalli che li separano, quello spaziale,

$$\sqrt{(x_1^{(1)}-x_1^{(2)})^2+(x_2^{(1)}-x_2^{(2)})^2+(x_3^{(1)}-x_3^{(2)})^2},$$

e quello temporale, $t^{(1)}-t^{(2)}$. In un altro sistema di coordinate, fisso rispetto al primo, entrambi gl'intervalli fra i punti restano gli stessi (si suppone che le unità di misura delle lunghezze e del tempo siano le medesime). Perciò una qualsiasi altra carta di questo tipo si ottiene dalla prima mediante una traslazione dell'origine e una rotazione rigida in $\mathbf{R}^3$ (lo spazio) e una traslazione dell'origine in $\mathbf{R}$ (il tempo). La presenza di un moto rettilineo uniforme del secondo sistema del laboratorio rispetto al primo conduce ad aggiungere alle coordinate in $\mathbf{R}^3$ ancora una traslazione lineare nel tempo. Cosí sono definite le «carte galileiane» e le «trasformazioni galileiane delle coordinate» dello spazio-tempo classici.

Spazio-tempo nella teoria della relatività ristretta. In questa teoria pure s'individua e si definisce una classe di carte, ognuna delle quali viene data mediante l'astrazione fisica di «sistema inerziale di coordinate». Approssimativamente è inerziale ogni sistema di coordinate del laboratorio connesso con un punto sulla superficie terrestre; un'approssimazione ancora migliore è fornita dai sistemi di coordinate connessi con il Sole o con una «stella fissa». Due qualsiasi sistemi inerziali di coordinate si muovono, uno rispetto all'altro, di moto rettilineo uniforme. Ogni sistema inerziale di coordinate identifica lo spazio-tempo con $\mathbf{R}^3 \times \mathbf{R}$, e il passaggio da un sistema a un altro è dato da funzioni lineari; queste, però, nel caso di sistemi in moto tra loro, non conservano l'asse dei tempi. Piú precisamente, se t è il tempo in un sistema e t' è il tempo in un altro sistema in moto rispetto al primo, una sezione «di tipo spaziale» $t = t_0$ dello spazio-tempo può non essere,

nell'altro sistema, rappresentata nella forma $t' = t'_0$. Nella teoria della relatività ristretta si postula che la velocità c della luce sia la stessa in tutti i sistemi inerziali, cosí che hanno significato invariante i «coni-luce» $x_1^2+x_2^2+x_3^2-c^2t^2=0$. Le trasformazioni per passare da un sistema all'altro conservano l'intervallo «spazio-temporale» fra i punti di M: $\Delta s^2 = \Delta x_1^2+\Delta x_2^2+\Delta x_3^2-c^2\Delta t^2$, dove $\Delta x_i = x_i - x'_i$, $\Delta t = t - t'$, (x, t) e (x', t') sono punti di M in un qualsiasi sistema inerziale di coordinate. Queste condizioni definiscono le equazioni di transizione da un sistema all'altro, che si chiamano trasformazioni di Lorentz.

Lo spazio $\mathbf{R}^4$ con la metrica definita da Δs^2 si chiama spazio di Minkowski. Gli spazi di Galileo e di Minkowski sono «piatti» nel senso della geometria differenziale.

Spazio-tempo nella teoria generale della relatività. In questa classe di modelli lo spazio-tempo M è rappresentato da una varietà differenziabile a quattro dimensioni, ricoperta da carte locali e dotata di una metrica locale che solo in un'approssimazione «infinitesimale» coincide localmente con la metrica di Minkowski. La forma generale del quadrato della distanza tra punti vicini (x_i) e (x_i+dx_i), $i = 1 \dots 4$, è dato da una forma quadratica

$$\sum_{i,k=1}^{4} g_{ik}\, dx_i\, dx_k,$$

dove le g_{ik} sono funzioni differenziabili delle coordinate locali e si chiamano «componenti del tensore metrico». Localmente questa forma può essere ridotta alla forma $y_1^2+y_2^2+y_3^2-y_4^2$, ciò che riflette la coincidenza «in piccolo» con la metrica di Minkowski. In termini di g_{ik} sono definite le grandezze che caratterizzano la «curvatura» dello spazio-tempo, e per queste grandezze si scrivono le equazioni differenziali di Einstein, le quali legano questa curvatura con la densità e il carattere del moto della materia-energia che si evolve nello spazio-tempo. La curvatura è localmente tanto maggiore quanto maggiore è la densità di materia-energia, il che spesso si esprime qualitativamente con l'affermazione: la gravitazione incurva lo spazio-tempo. Questa curvatura si esprime nel fatto, per esempio, che l'area della sfera con raggio crescente e centro in un dato punto di M varia con una legge diversa da quella per lo spazio-tempo piatto.

Nei modelli cosmologici di M spesso si suppone che la materia e l'energia siano distribuite in M uniformemente e si muovano nello spazio isotropicamente; non vi sono direzioni privilegiate. Benché ciò contraddica i dati osservabili fino alle dimensioni galattiche, si presume che a dimensioni notevolmente maggiori la distribuzione discreta della materia, dopo una conveniente media, possa essere trascurata, di modo che in tali modelli globali venga a dominare l'aspetto continuo.

Spazio delle fasi. Nello spazio-tempo M il moto di un punto materiale è dato dalla sua traiettoria, una curva in M, che si chiama linea di universo del punto. Ad eccezione dei modelli cosmologici, che considerano M globalmente, oggetti della fisica sono i «sistemi isolati», cioè le regioni di M, e la materia in queste regioni, che si evolve in modo tale che l'influenza su di esse delle altre parti

di M possa essere, in un'approssimazione nota, trascurata. Spesso in questi casi un modello matematico del sistema è il suo spazio delle fasi N, unitamente alle curve o traiettorie su esso, che formano il quadro delle fasi del sistema.

I punti dello spazio delle fasi N corrispondono a tutti i possibili stati istantanei del sistema. Le curve delle fasi su N corrispondono a tutti i possibili percorsi dell'evoluzione del sistema. Molto spesso N è dotato in modo naturale di una struttura di varietà differenziabile e le curve delle fasi sono soluzioni di equazioni differenziali.

Per esempio, si supponga che in un dominio U dello spazio galileiano possano muoversi k punti materiali sotto l'influenza delle forze di mutua interazione (ad esempio gravitazionale) e, eventualmente, per effetto di altre forze (per esempio, di un campo di forze esterno). In accordo con i principî della meccanica di Newton l'evoluzione di questo sistema è definita dalle posizioni iniziali e dalle velocità iniziali dei punti, cioè in tutto da $6k$ numeri (3 coordinate cartesiane e 3 componenti del vettore velocità per ogni punto). Lo spazio delle fasi N di questo sistema sarà una varietà a $6k$ dimensioni U^k (posizioni) $\times \mathbf{R}^{3k}$ (velocità), il secondo fattore essendo il fibrato tangente a U^k (cfr. ancora l'articolo «Applicazioni»). Lungo ogni curva delle fasi sono soddisfatte equazioni differenziali del tipo $dq_i^{(j)}/dt = p_i^{(j)}$, $dp_i^{(j)}/dt = F(p, q, t)$, dove le $q_i^{(j)}$ sono le coordinate del punto i-esimo; le $p_i^{(j)}$ sono le componenti della sua velocità, $p = (p_i^{(j)})$, $q = (q_i^{(j)})$; t è il tempo. La funzione F è, insieme con lo spazio delle fasi, la caratteristica matematica fondamentale del sistema.

Un'analisi dettagliata permette d'individuare una classe molto importante di sistemi, nei quali gli spazi delle fasi e le equazioni del moto hanno una struttura speciale: i cosiddetti sistemi hamiltoniani.

Sia N lo spazio delle fasi di un certo sistema e si chiamino osservabili (macroscopiche) del sistema le funzioni differenziabili su N. Le osservabili corrispondono alle diverse grandezze fisiche che possono essere misurate in tutti i possibili stati del sistema.

Il sistema si dice hamiltoniano se sono soddisfatte le seguenti condizioni:

I) Sull'insieme delle osservabili è data una legge binaria di composizione $F, G \to \{F, G\}$ (cfr. l'articolo «Applicazioni») che si chiama parentesi di Poisson e definisce sulle osservabili la struttura di algebra di Lie. Con ciò, per ogni punto $x \in N$ esistono le osservabili $q_1, \ldots, q_r$ («coordinate generalizzate») e $p_1, \ldots, p_r$ («impulsi generalizzati») tali che in un intorno del punto x esse sono le coordinate locali e, in questo intorno, la parentesi di Poisson è data dall'uguaglianza

$$\{F, G\} = \sum_{i=1}^{r} \left(\frac{\partial F}{\partial p_i} \frac{\partial G}{\partial q_i} - \frac{\partial G}{\partial p_i} \frac{\partial F}{\partial q_i} \right).$$

Tali coordinate (q, p) si chiamano canoniche; esse non sono definite univocamente.

II) È data un'osservabile H, la hamiltoniana del sistema, o l'energia osservabile, tale che su ogni traiettoria, che si evolva nel tempo, ogni osservabile F soddisfa l'equazione differenziale $dF/dt = \{H, F\}$.

Le coordinate canoniche (q, p) localmente formano un sistema completo di osservabili, attraverso le quali si esprimono tutte le altre. Perciò la condizione II) è equivalente al sistema di equazioni differenziali di Hamilton

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \quad \frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}.$$

Assegnando H e la posizione iniziale del sistema e risolvendo questo sistema di equazioni, si ottiene una traiettoria concreta del sistema che descrive la sua evoluzione.

Si consideri, per esempio, un sistema di k punti materiali che si attraggono secondo la legge di Newton. Il suo spazio delle fasi $\mathbf{R}^{3k}$(coordinate) $\times \mathbf{R}^{3k}$(impulsi) è il supporto del sistema hamiltoniano. Nello spazio vi è il sistema canonico globale di coordinate: $q_i^{(j)}$ = coordinate cartesiane dei punti ($j = 1, 2, 3$; $i = 1, \ldots, k$); $p_i^{(j)}$ = componenti degli impulsi m_i $(dq_i^{(j)}/dt)$ sugli assi corrispondenti (m_i sono le masse). La hamiltoniana ha la forma

$$\sum_{i=1}^{k} \frac{p_i^2}{2m_i} - G \sum_{i,j=1}^{k} \frac{m_i m_j}{r_{ij}},$$

dove r_{ij} è la distanza tra i-esimo e j-esimo punto, $p_i^2 = p_i^{(1)^2} + p_i^{(2)^2} + p_i^{(3)^2}$. Si osservi che qui la hamiltoniana ha delle singolarità per $r_{ij} = 0$ («urti»).

Le equazioni differenziali descrivono l'evoluzione anche di molti altri sistemi in un'enorme classe di modelli, da quelli fisici a quelli ecologici. In queste descrizioni a priori domina l'aspetto continuo. Si mostrerà ora attraverso alcune tipiche nozioni come appaiano le caratteristiche discrete. Esse si riferiscono ai tipi di evoluzione del sistema ed è comodo descriverle in termini del suo quadro delle fasi, cioè della disposizione delle traiettorie nello spazio delle fasi. L'«individualità» di ogni storia concreta del sistema è descritta in primo luogo dalle caratteristiche topologiche discrete nel continuo di tutte le traiettorie.

Stati di equilibrio. Sono punti dello spazio delle fasi nei quali le velocità sono nulle. Da un tale stato iniziale il sistema non esce in alcun modo. Nel caso hamiltoniano si tratta dei punti dove $\partial H/\partial p_i = \partial H/\partial q_i = 0$.

Il caso piú generale è dato dalle varietà invarianti nello spazio delle fasi, cioè quelle varietà tali che, se il punto vi capita nel momento iniziale, resta eternamente in esse. Sono varietà invarianti le linee di livello degl'integrali del sistema, cioè le osservabili i cui valori non cambiano nel corso dell'evoluzione. L'esempio piú importante d'integrale di un sistema hamiltoniano conservativo è la sua energia, cioè la hamiltoniana H.

Un sistema con uno spazio delle fasi a $2n$ dimensioni si dice completamente integrabile se si può fibrare in varietà invarianti a n dimensioni, su ognuna delle quali il moto proceda, in convenienti coordinate locali, linearmente. Quando il sistema è hamiltoniano e queste varietà sono compatte, esse sono dei tori a n dimensioni e su esse il moto è quasi periodico: in adeguate coordinate angolari, $\varphi_1, \ldots, \varphi_n$, definite mod 2π, il moto è descritto dalle equazioni $\varphi_i(t) - \varphi_i(0) = \omega_i t$, dove le ω_i sono le frequenze del moto sul toro; $\varphi_i(0)$ sono i valori iniziali delle

coordinate. Solitamente tali tori sono individuati dalle equazioni $I_1 = \text{cost}, \ldots, I_n = \text{cost}$, dove gli I_j sono integrali speciali del sistema, detti anche azioni variabili. Il moto su ogni toro può essere immaginato come un insieme di n oscillatori che oscillano concordemente con le frequenze ω_i.

Stabilità e attrattori. Non tutti gli stati di equilibrio descrivono la storia reale del sistema. Le condizioni iniziali di qualsiasi sistema sono note solo approssimativamente e può accadere che tutte le traiettorie che cominciano vicino allo stato di equilibrio si allontanino da esso (esempio: un pendolo costituito da una sbarra rigida nella posizione superiore). Tale stato di equilibrio si dice instabile. Al contrario, se condizioni iniziali vicine allo stato di equilibrio dànno traiettorie che restano vicine a quella di equilibrio, tale stato è stabile (con significati diversi, a seconda della scelta della definizione matematica). Analogamente, le varietà invarianti possono essere stabili o instabili. Per esempio, sul piano delle fasi può esservi una traiettoria chiusa senza intersezioni. Le traiettorie vicine ad essa possono avvolgervisi attorno da entrambi i lati (ciclo limite stabile), o da un lato solo (ciclo limite semistabile), oppure svolgersi da entrambi i lati (instabilità). Esse possono comportarsi anche in altri modi; per esempio, un anello attorno alla traiettoria chiusa può essere riempito da traiettorie chiuse parallele. Gli insiemi, del tipo ciclo limite stabile, verso i quali si avvicinano tutte le traiettorie che comincino vicino ad essi, si chiamano attrattori. Ogni attrattore ha il suo dominio di attrazione. Questi dominî sono divisi dalle varietà separatrici. La configurazione degli attrattori e delle varietà separatrici, quando esistono, sono caratteristiche discrete essenziali del sistema: esse definiscono i regimi limite e il loro andamento. Molti sistemi elettrici e meccanici, in particolare generatori di oscillazioni, sono costruiti in modo tale che nell'intorno del loro regime caratteristico di funzionamento esistano questi regimi limite.

Una classe importante di effetti discreti è legata alla possibilità che il sistema abbia delle instabilità che si esprimano bruscamente, il cui tipo di comportamento, a lungo andare, in generale può dipendere in modo discreto dai dati iniziali.

Per esempio, l'esistenza del sistema solare come un tutt'unico significa che il Sole e i pianeti si trovano in uno «stato legato»: le loro mutue distanze sono limitate. È stabile questo stato? Cioè, non potrebbe accadere che un pianeta sia «espulso» e se ne vada verso l'infinito?

Il calcolo approssimato del moto di un pianeta con i metodi della teoria delle perturbazioni implica l'uso delle serie. Nei denominatori dei loro coefficienti figurano combinazioni lineari della forma $k_1\omega_1 + \ldots + k_n\omega_n$, dove i k_i sono numeri interi e le ω_i sono le frequenze di rivoluzione. Il meccanismo che conduce a questo risultato qualitativamente può essere chiarito cosí: in una prima approssimazione il moto di ciascun pianeta può essere calcolato supponendo che gli altri pianeti siano diffusi uniformemente lungo le rispettive orbite, di modo che il campo gravitazionale da essi prodotto resti invariato. Però questa approssimazione si scosta tanto piú fortemente dalla realtà quanto piú spesso i pianeti, nel loro moto, si raccolgono in gruppi che si trovino su uno stesso raggio. Se le frequenze di rivoluzione sono linearmente dipendenti, cioè, se $k_1^0\omega_1 + \ldots + k_n^0\omega_n = 0$, una tale posizione si ripete infinite volte. Se una tale combinazione lineare con k_i^0 non grandi, pur non essendo uguale a zero, risulta molto piccola, s'introdurrà lo stesso una notevole perturbazione. Per esempio, Giove e Saturno in un giorno percorrono circa 299″ e 120,5″ rispettivamente, cosí che il denominatore $2\omega_G - 5\omega_S$ è molto piccolo. Ciò produce una grande perturbazione dei pianeti con un periodo di circa ottocento anni. L'accumulazione di tali perturbazioni in un sistema tipo quello planetario può condurre alla sua disintegrazione. Pertanto, in questa idealizzazione, la storia di un sistema meccanico può dipendere essenzialmente da proprietà aritmetiche dello stato iniziale, quali la commensurabilità delle frequenze di rivoluzione iniziali. Un'esatta indagine matematica delle traiettorie di fase di tali sistemi permette di chiarire la struttura straordinariamente complessa del loro quadro delle fasi (teoria di Kolmogorov-Arnol'd-Moser). Per esempio, se la hamiltoniana del sistema si discosta poco dalla hamiltoniana di un sistema integrabile («perturbazione debole»), si verifica che, sotto certe condizioni, vi è un insieme di tori invarianti del sistema non perturbato, avente misura di Lebesgue grande, che non sparisce, e che solo alcuni tori si deformano e il moto su questi resta condizionatamente periodico. Tuttavia, per $n > 2$, le traiettorie che non capitano su tori deformati possono vagare tra i loro interstizi, che hanno struttura complessa, e allontanarsi quanto si vuole dallo stato iniziale.

Un andamento ancor piú caotico si rivela nei tipi di moto lungo traiettorie geodetiche su varietà con curvatura negativa. In questi problemi le questioni sulle traiettorie individuali perdono quasi completamente di significato e il modo naturale di porre i problemi è in una certa misura «complementare»: si svolgono indagini sulle caratteristiche continue del moto che si ottengono mediando le traiettorie sullo spazio delle fasi e sul tempo. È, questo, l'oggetto della meccanica statistica (cfr. oltre).

Controllo e catastrofi. Un aspetto importante del comportamento «discreto» di un sistema continuo, che risale ai lavori di Poincaré, è stato sottolineato e studiato a fondo da Thom [1971] e dalla sua scuola e ha preso il nome di «teoria delle catastrofi». Nella teoria delle catastrofi si considera una famiglia di sistemi dipendenti da alcuni «parametri di controllo». Sia $c \in C$ un punto nello spazio dei parametri di controllo, N_c sia il corrispondente spazio delle fasi. Variando c, la configurazione degli attrattori e delle varietà separatrici del quadro delle fasi di N_c può cambiare; per esempio, possono sparire e/o apparire stati di equilibrio; stati di equilibrio stabili possono diventare instabili, ecc. Nello stesso tempo la variazione del quadro delle fasi conserva il suo tipo topologico fino a quando c non interseca alcuni punti critici nello spazio dei parametri di controllo C e in questi punti la variazione avviene con un salto. Si supponga ora che il sistema controllato si sia evoluto lungo traiettorie vicine a un attrattore stabile A_c e che i parametri di controllo c siano variati lentamente e, raggiungendo il valore c_0, l'attrattore A_c scompaia, o diventi instabile. Allora il sistema, di regola, cambia repentinamente il tipo di comportamento per portarsi su un altro regime limite.

Si consideri, per esempio, la seguente illustrazione meccanica. Il sistema con-

trollabile sia una sfera che rotola lungo una curva piana (una scanalatura) sotto l'azione della forza di gravità. I minimi locali di questa curva sono stati di equilibrio stabili; per semplicità si riterrà che l'attrito sia cosí grande che da ogni posizione la sfera rotoli verso il minimo piú vicino e qui essa si arresti senza ulteriori oscillazioni. Si supponga poi che i parametri di controllo cambino la forma della curva. Al momento iniziale la sfera si trovi in un certo minimo; sulla curva si formi poi un altro minimo contiguo al precedente, ma piú basso, e quindi la barriera tra i due minimi cominci ad abbassarsi. Quando questa giunge al livello del primo minimo, quest'ultimo diventa instabile e in un istante successivo la sfera rotola nel nuovo minimo.

Tali passaggi di un sistema da un regime di comportamento a un altro, sotto forma di salti in seguito a una variazione lenta di alcuni parametri, sono stati studiati da Thom in varie situazioni biologiche e sociali.

Meccanica statistica dell'equilibrio. Si consideri un volume macroscopico, per esempio una mole di gas a una data temperatura. Se è sufficientemente rarefatto, il suo comportamento può essere modellato riguardando le molecole quali particelle classiche in moto secondo le leggi della meccanica di Newton e interagenti tra loro mediante un certo potenziale di repulsione. Poiché una mole di gas contiene circa $6 \cdot 10^{23}$ molecole, lo spazio delle fasi del corrispondente sistema meccanico ha una dimensione enorme. L'energia E è un integrale del moto, ma l'iperpiano $E = \text{cost}$ ha una dimensione inferiore solo di uno. Le curve delle fasi su di esso formano un flusso che conserva la misura di Liouville $dp_1 \ldots dp_{3k} dq_1 \ldots dq_{3k}$ (si veda l'articolo «Applicazioni» di questa stessa *Enciclopedia*). Il comportamento individuale delle curve delle fasi è eccezionalmente complesso e non è possibile indagarlo. Tuttavia nello studio macroscopico degli stati di equilibrio ci si interessa a un piccolo numero di osservabili quali la pressione, la densità, ed esse mostrano un comportamento di una regolarità e semplicità sorprendenti. Secondo Maxwell, Boltzmann e Gibbs ciò è dovuto al fatto che queste variabili macroscopiche si ottengono mediando le caratteristiche microscopiche e il risultato teorico di questa media, pur non coincidendo necessariamente con l'osservabile, è vicino ad esso con una probabilità quasi uguale all'unità.

Si esamineranno piú da vicino le procedure di mediazione qui essenziali. Sia f una funzione su N. Il valore di questa funzione in ogni punto – stato del sistema – è il valore di una certa osservabile macroscopica. Si consideri un'evoluzione del sistema e sia $x(t) \in N$ il punto corrispondente dello spazio delle fasi all'istante t. L'esperienza mostra che il valore di $f(x(t))$ ha un limite. Esso può essere espresso nel modo seguente:

$$\lim_{T \to +\infty} \frac{1}{T} \int_0^T f(x(t))\, dt = \bar{f}(x),$$

dove $x = x(0)$. La funzione $\bar{f}$ è una media temporale lungo una data traiettoria ed essa ha un immediato significato fisico. Però calcolarla è difficile data la complessità del comportamento individuale di una traiettoria. Si prestano al calcolo, invece, alcune medie spaziali, cioè degl'integrali estesi allo spazio delle fasi (piú

precisamente, estesi a $E = H = \text{cost}$) con la misura di Liouville. Il teorema di Birkhoff afferma che, se sono soddisfatte alcune condizioni di ergodicità, le medie temporali esistono e coincidono con le medie spaziali tranne che per un insieme di condizioni iniziali di misura nulla. La causa intuitiva di ciò sta nel fatto che quasi tutte le traiettorie individuali, dopo un lungo tempo, capitano nell'intorno di ogni punto dello spazio delle fasi e vi si attorcigliano attorno con una densità proporzionale alla misura di Liouville.

Meccanica statistica del non-equilibrio. I suoi modelli hanno a che fare con processi quali il comportamento di un certo volume di gas, precedentemente suddiviso in due metà a pressione e temperatura diverse, dopo aver tolto la parete divisoria, oppure la condensazione di un gas in un liquido, o il raffreddamento di un liquido in un solido per abbassamento della temperatura, cioè transizioni di fase.

Nei problemi del primo tipo sorge il problema del passaggio dalle equazioni hamiltoniane del moto delle molecole alle equazioni mediate dell'idrodinamica, che descrivono il campo delle velocità, i processi di trasporto del calore, ecc. Queste equazioni differenziali non lineari alle derivate parziali corrispondono a effetti «discreti» quali le onde d'urto.

Nei problemi del secondo tipo sorge il problema di trovare i valori critici di quelle grandezze macroscopiche per i quali avvengono salti nella densità, nella capacità termica, ecc., e appare la struttura «granulare» del mezzo, cioè la frontiera tra le fasi.

Una delle difficoltà di principio della fisica statistica è capire in qual modo la reversibilità del tempo nelle equazioni del moto delle particelle si accordi con la sua irreversibilità nel comportamento macroscopico del mezzo. Quest'ultima si riflette nel fatto che il calore passa dalle regioni con temperatura piú alta alle regioni con temperatura piú bassa e non viceversa; la viscosità rallenta il flusso dei liquidi e dei gas e cosí via. Una spiegazione qualitativa di questi effetti di nuovo consiste nel fatto che queste leggi macroscopiche sono soddisfatte solo con una grande probabilità. Tuttavia le difficoltà del calcolo sono molto grandi e una teoria conseguente della meccanica statistica del non-equilibrio fino ad ora non esiste.

Onde. Nella teoria macroscopica dei mezzi continui hanno una importanza eccezionale i fenomeni ondulatori. I loro modelli matematici – con una nuova interpretazione – hanno parimenti un'importanza fondamentale nella meccanica quantistica.

Il piú semplice fenomeno ondulatorio è dato dalle oscillazioni armoniche di una certa grandezza con frequenza ω. La dipendenza di questa grandezza dal tempo è della forma $Ae^{i\omega t}$, dove ω è un numero reale. Il loro periodo è uguale a $2\pi/\omega$, cosicché ω è la frequenza circolare delle oscillazioni. (Se la grandezza è reale, si può scrivere $A \cos \omega(t - t_0)$, ma le notazioni complesse sono comode matematicamente, e nella meccanica quantistica gli oggetti fondamentali oscillanti – le funzioni di stato ψ – in generale sono complesse). Per esempio, oscilla

in questo modo lo spostamento di un punto materiale su una molla finché la forza elastica della molla è soggetta alla legge di Hooke, cioè è proporzionale all'allungamento della molla. In questo modo oscilla l'intensità della corrente nel piú semplice circuito oscillante. Le oscillazioni di qualsiasi forma con frequenza ω, che si ripetono cioè rigorosamente dopo $2\pi/\omega$ unità di tempo, possono essere rappresentate da una serie di Fourier

$$\sum_{n=-\infty}^{+\infty} A_n e^{in\omega t},$$

cioè le si può ottenere mediante una sovrapposizione di oscillazioni lineari. Se al posto della serie di Fourier si scrive l'integrale di Fourier $F(t)=\int_{-\infty}^{+\infty} A(\omega)e^{i\omega t}d\omega$, dove $A(\omega)$ è una funzione delle frequenze, si possono ottenere diverse leggi di variazione di una grandezza rispetto al tempo, espresse sotto forma di sovrapposizione di oscillazioni armoniche, dove però lo spettro di frequenze è ora continuo. Tra queste leggi vi sono quelle rappresentate da «picchi» della grandezza $F(t)$ arbitrariamente grandi e stretti per esempio all'istante $t=0$ e con valori vicini a zero in tutti gli altri istanti, e ciò a causa dell'interferenza fra le onde.

Se la grandezza ha carattere di campo, cioè dipende anche, per esempio, da una coordinata spaziale x, è possibile considerare l'onda armonica $Ae^{i(kx+\omega t)}$. Qui k è il numero d'onda: $2\pi/k$ è il periodo spaziale dell'onda, mentre $2\pi/\omega$ è quello temporale. Muovendosi lungo l'asse x con velocità $-\omega/k$, cioè passando al sistema di coordinate $x'=x+(\omega/k)t$, si ottiene l'onda $Ae^{ikx'}$ indipendente dal tempo; in altre parole, l'onda armonica considerata si muove uniformemente nello spazio con velocità $-\omega/k$ senza cambiare la sua forma. Nei mezzi lineari le onde armoniche si muovono tutte con una stessa velocità v e un moto ondulatorio generico si ottiene quale sovrapposizione di queste onde armoniche con k diversi, con $\omega=vk$ e con le ampiezze $A(k)$. Un'opportuna sovrapposizione $F(x, t)=\int_{-\infty}^{+\infty} A(k)e^{ik(x-vt)}dk$ si può rappresentare come un picco stretto vicino al punto $x=vt$ che scende rapidamente a zero. Esso si muove con velocità v come un tutt'unico. Queste formazioni si chiamano «pacchetti d'onda». Per molti versi essi sono simili a «particelle discrete» e si ottengono quale risultato dell'interferenza fra «onde continue». Si osservi che si può ottenere un pacchetto d'onda ben localizzato nello spazio-tempo solo quale somma di onde con numeri d'onda, o frequenze, fortemente diffusi. Inversamente un'onda con valori puntuali di k, ω è «diffusa» nello spazio-tempo. Un'esatta formulazione di queste relazioni è il fondamento del principio d'indeterminazione di Heisenberg in meccanica quantistica, dove esse vengono applicate alla funzione d'onda ψ.

In certi mezzi possono pure propagarsi onde armoniche, ma con altre relazioni fra il numero d'onda k e la frequenza ω; in particolare, le velocità di queste onde possono essere diverse. Inoltre, nei mezzi non-lineari, in generale, non si può realizzare la somma di due perturbazioni che si propagano. Ciò nonostante sono noti mezzi non-lineari assai interessanti, nei quali possono diffondersi perturbazioni ondulatorie localizzate, che si comportano come un tutt'unico e che asintoticamente conservano la propria forma perfino dopo urti e interazioni. Tali formazioni si chiamano onde isolate o «solitoni». I loro primi esempi furono

scoperti da Boussinesq, Korteweg e De Vries indagando la propagazione di onde lunghe sulla superficie di acque non profonde in una certa approssimazione non-lineare. In questi ultimi anni sono state scoperte moltissime equazioni d'onda non-lineari aventi tali soluzioni solitoniche e multisolitoniche. Vi è la speranza di scoprire equazioni d'onda realistiche della meccanica quantistica, dove tali soluzioni descrivano particelle elementari.

Si torni ora al caso di un «oscillatore armonico», cioè di una grandezza che oscilla nel tempo come $Ae^{i\omega_0 t}$. Questo processo corrisponde a oscillazioni «libere» e avviene senza scambio di energia con l'esterno. Si supponga ora che su tale grandezza agisca una perturbazione esterna periodica con frequenza ω. Per esempio, una sferetta su una molla fatta oscillare con una forza esterna. Allora l'ampiezza di tali oscillazioni forzate, a mano a mano che ω si approssima a ω_0, crescerà (in un modello idealizzato fino a ∞); cioè si verificherà una risonanza. Corrispondentemente, il grafico della quantità d'energia che l'oscillatore, in queste condizioni, è in grado di assorbire, in dipendenza della frequenza ω della perturbazione esterna, ha un picco nel punto $\omega=\omega_0$. Pertanto le risonanze possono essere riguardate quali fenomeni di tipo discreto nello spazio frequenze-energie.

Campo elettromagnetico. Le cariche elettriche connesse con la materia nello spazio-tempo generano in esso un campo elettromagnetico. Nei modelli della teoria della relatività ristretta, fissato un sistema inerziale di coordinate, tale campo è dato da due vettori spaziali, **E** e **H**, definiti in ogni punto dello spazio-tempo. Il vettore **E** – intensità del campo elettrico – definisce la forza che agisce sulla carica elettrica unitaria in quiete in quel punto. Il vettore **H** – intensità del campo magnetico – definisce la componente della forza che agisce sull'unità elettrica in moto, e questa componente è proporzionale alla velocità del moto e perpendicolare ad essa. I vettori **E** ed **H** dipendono dai punti dello spazio-tempo, sono differenziabili e soddisfano un sistema di equazioni differenziali alle derivate parziali: le equazioni di Maxwell. La densità dell'energia trasportata dal campo è proporzionale al quadrato dell'ampiezza, cioè a E^2+H^2.

Anche in questa teoria, pertanto, domina l'aspetto continuo. Il discreto appare solo nella descrizione del moto delle stesse cariche elettriche che generano il campo e che interagiscono con esso e fra loro. Tuttavia nell'elettrodinamica classica non vi è posto per i fotoni e non è possibile dare una spiegazione ai fenomeni nei quali le proprietà quantistiche della radiazione elettromagnetica svolgono un ruolo essenziale.

Infatti, nell'ambito della teoria classica, un'onda elettromagnetica con frequenza ω, che si propaga da una sorgente puntiforme, è a simmetria sferica; la densità di energia da essa trasportata diventa arbitrariamente piccola allontanandosi dalla sorgente e non sono possibili sue improvvise concentrazioni di quanti $h\omega/2\pi$ in fenomeni quali, per esempio, gli effetti fotoelettrici.

Nella teoria classica un aspetto discreto interviene dall'esterno, considerando il campo e le cariche che lo producono separatamente. Una carica puntiforme e produce attorno a sé un campo elettrico statico con un'intensità e/r^2 alla distanza r dalla carica stessa. Pertanto essa si presenta quale singolarità del campo, cioè

come punto nel quale il campo diventa infinito (cfr. l'articolo «Applicazioni» di questa stessa *Enciclopedia*).

Rappresentazioni analoghe sono usate nella teoria del movimento della materia nell'ambito della teoria generale della relatività, dove gli addensamenti di materia si possono riguardare quali singolarità del campo gravitazionale e le leggi del moto possono dedursi dalle equazioni del campo.

1.3. Modelli teorici della microfisica.

Lo spazio-tempo nei modelli fondamentali della microfisica è un certo dominio nello spazio-tempo galileiano (teoria non-relativistica), oppure nello spazio-tempo di Minkowski (teoria relativistica). La teoria non-relativistica ha a che fare con un'approssimazione, nella quale il moto avviene con una velocità piccola rispetto a quella della luce e perciò questa viene ritenuta infinita. L'ordine di questa piccolezza definisce il carattere dell'approssimazione. Per esempio, nel modello semiclassico dell'atomo d'idrogeno la velocità media dell'elettrone è circa uguale a $c/137$, ciò che giustifica l'approssimazione non-relativistica. Le velocità sono legate ai processi energetici. Perciò i processi ad alte energie devono essere studiati nella teoria relativistica cosí come gli «effetti fini» quali, per esempio, la separazione delle linee spettrali in processi energetici relativamente deboli.

Sovente gli spazi delle fasi di sistemi isolati appaiono nei modelli della meccanica quantistica come oggetti intermedi dell'indagine; inizialmente il sistema viene descritto mediante il suo spazio delle fasi classico con le relative osservabili; dopo di che vi si applicano speciali «regole di quantizzazione». Si descriveranno ora nei particolari alcuni esempi.

Nella meccanica quantistica, cosí come in quella classica, inizialmente si descrive la situazione in termini di osservabili e di stati, tuttavia le une e gli altri corrispondono a oggetti matematici di natura diversa.

Gli stati (o stati puri) di un sistema meccanico-quantistico sono raggi, o spazi a una dimensione in un certo spazio di Hilbert $\mathcal{H}$, che è uno spazio lineare sul corpo dei numeri complessi $\mathbf{C}$, nel quale è dato un prodotto scalare (f, g) con le proprietà $(af, g) = a(f, g)$ e $(f, g) = (\bar{g}, \bar{f})$ ($\bar{a}$ è il numero complesso coniugato di a). Posto $\|f\| = (f, f)^{1/2}$, si chiede che da $\|f\| = 0$ segua che $f = 0$. Questa norma definisce su $\mathcal{H}$ una topologia e lo spazio $\mathcal{H}$ dev'essere completo in questa topologia, cioè ogni successione $\{f_n\}$ con $\|f_n - f_m\| \to 0$ per $\min(n, m) \to \infty$ converge a un certo elemento di $\mathcal{H}$.

Gli spazi tipici $\mathcal{H}$ che appaiono nella meccanica quantistica contengono tutte le funzioni differenziabili f su una certa varietà N con misura μ, per le quali $\int_N |f|^2 d\mu < \infty$, e cosí pure tutte le «funzioni» limite rispetto al prodotto scalare $(f, g) = \int_N f\bar{g}\, d\mu$. Questi spazi si indicano con $L_2(N)$.

In particolare, se un sistema classico hamiltoniano è descritto mediante le coordinate canoniche globali $q_1, \ldots, q_n, p_1, \ldots, p_n$, le funzioni d'onda o di stato $\psi(q_1, \ldots, q_n)$ con $\|\psi\|^2 = \int |\psi|^2\, dq_1 \ldots dq_n$ appartengono allo spazio di Hilbert corrispondente a questo sistema e descrivono stati possibili del sistema nella teoria quantistica.

Assegnare lo stato quantistico di un sistema mediante la funzione ψ in questa rappresentazione delle «coordinate» permette di ottenere su esso l'informazione seguente: la probabilità che, in un esperimento per misurare le coordinate generalizzate $q_1, \ldots, q_n$, i loro valori giacciano in un dominio U è uguale a $\int_U |\psi|^2\, dq_1 \ldots dq_n / \|\psi\|^2$.

Pertanto in prima approssimazione la descrizione quantistica degli stati di un sistema introduce un nuovo aspetto della continuità (è la funzione di stato o d'onda ψ «diffusa» invece dei punti nello spazio delle coordinate-impulsi) e una probabilità che non figura nella descrizione classica. Una particella ben localizzata corrisponde a un «pacchetto d'onda» di funzioni ψ.

Passando ora alle osservabili, si rileva come in meccanica quantistica a ogni osservabile corrisponda un operatore lineare A che agisce sullo spazio degli stati $\mathcal{H}$ del sistema.

Alle osservabili a valori reali corrispondono operatori autoaggiunti, per i quali $(Af, g) = (f, Ag)$ per tutti gli $f, g \in \mathcal{H}$. Se $f \in \mathcal{H}$ è tale che $Af = \lambda f$ per un certo numero λ, allora questo numero si chiama autovalore di A e f si dice autovettore corrispondente a questo autovalore. Gli autovalori di un operatore autoaggiunto A sono sempre reali; essi formano una parte dello spettro dell'operatore A, che, per definizione, è formato da quei λ per i quali $A - \lambda I$ non è invertibile (I è l'operatore identità). Nel caso in cui A è autoaggiunto e soddisfa la condizione supplementare di completa continuità, il suo spettro è costituito esattamente dai suoi autovalori che sono numerabili, mentre con gli autovettori di A si può formare una base ortonormale dello spazio $\mathcal{H}$. Si osservi che a uno stesso autovalore λ può corrispondere un intero spazio vettoriale di autovettori; in fisica la sua dimensione si chiama grado di degenerazione di λ. Nel caso generale nello spettro di A, oltre la componente discreta, vi può essere una componente continua.

Sia A una certa osservabile di un sistema quantistico con spazio degli stati $\mathcal{H}$. Sia $\psi \in \mathcal{H}$ un vettore di stato del sistema. Si supponga per semplicità che $\psi_1, \ldots, \psi_n, \ldots$ sia una base ortonormale in $\mathcal{H}$, cioè $\|\psi_i\| = 1$, $(\psi_i, \psi_j) = 0$, $A\psi_i = \lambda_i \psi_i$, e si scomponga ψ in questa base:

$$\psi = \sum_{i=1}^{\infty} a_i \psi_i .$$

Allora, misurando l'osservabile A nello stato ψ, si possono ottenere solo i valori λ_i e ogni valore λ_i può essere ottenuto con la probabilità $|a_i|^2 / \|\psi\|^2$.

Questo è il modello matematico fondamentale mediante il quale appaiono le caratteristiche quantistiche discrete del sistema a partire dai metodi continui della sua descrizione. In presenza di una componente continua dello spettro i valori di un'osservabile in uno stato sono pure distribuiti in modo continuo con una certa densità di probabilità, ma qui una esatta descrizione del risultato è matematicamente piú complessa e non ci si soffermerà.

Le osservabili quantistiche possono essere di tipo classico e non-classico. Delle prime fanno parte l'energia, le coordinate generalizzate e gl'impulsi. Delle seconde fanno parte lo spin e cosí pure gli altri diversi numeri quantici.

Le regole fondamentali per ottenere le osservabili quantistiche di tipo classico sono le seguenti. Se $q_1, \ldots, q_n, p_1, \ldots, p_n$ sono le osservabili classiche delle coordinate generalizzate e degl'impulsi di un sistema hamiltoniano e gli elementi di $\mathcal{H}$ si considerano nella rappresentazione delle coordinate $\psi(q_1, \ldots, q_n)$, allora l'operatore della coordinata q_j è la moltiplicazione per q_j, mentre l'operatore dell'impulso p_j è $-ih(\partial/\partial q_j)$, dove h è la costante di Planck. Se la hamiltoniana del sistema classico ha la forma $V(q)+T(p)$ (come in un sistema di particelle con interazioni gravitazionali, o elettromagnetiche), il suo corrispondente operatore dell'energia H si ottiene sotto forma di somma dell'operatore di moltiplicazione per $V(q)$ e dell'operatore differenziale $T(-ih(\partial/\partial q_j))$. L'operatore dell'energia si chiama anche «operatore di Schrödinger» del sistema.

L'evoluzione del sistema può essere descritta in due modi: o quale evoluzione del suo vettore di stato in $\mathcal{H}$ con le osservabili costanti (rappresentazione di Schrödinger), oppure quale evoluzione degli operatori delle osservabili con il vettore di stato costante (rappresentazione di Heisenberg). Formalmente la seconda descrizione è completamente analoga a quella hamiltoniana: l'evoluzione dell'osservabile A soddisfa l'equazione differenziale

$$\frac{dA}{dt}=\left[\frac{i}{h}H, A\right],$$

dove al posto della parentesi di Poisson appare l'operazione di commutazione $[A, B]=AB-BA$, che introduce nell'insieme delle osservabili quantistiche la struttura di algebra di Lie. Nella prima descrizione si ottiene l'equazione di Schrödinger $ih(d\psi/dt)=H\psi$.

Da un punto di vista matematico si può dire che il passaggio dalle osservabili classiche d'un sistema hamiltoniano a quelle quantistiche consiste nella scelta della rappresentazione dell'algebra di Lie-Poisson delle osservabili classiche nell'algebra di Lie degli operatori autoaggiunti d'uno spazio di Hilbert. Questa rappresentazione conserva l'equazione dell'evoluzione (nelle forme di Hamilton e di Heisenberg).

Qui di seguito si dà una breve descrizione di come si usino questi concetti generali per spiegare fenomeni concreti del micromondo.

Livelli di energia. Lo stato non-relativistico dell'atomo d'idrogeno, in un sistema di coordinate rigidamente connesso con il nucleo e ove non si tenga conto dello spin dell'elettrone, è dato da una funzione ψ nello spazio fisico euclideo $\mathbf{R}^3$ con $\int|\psi|^2<\infty$. L'operatore dell'energia H ha la forma

$$H\psi=-\frac{h^2}{8\pi^2 m}\left(\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}\right)-\frac{e^2\psi}{\sqrt{x^2+y^2+z^2}},$$

dove e è la carica dell'elettrone e m è la sua massa. Uno studio analitico mostra che questo operatore è dotato di una successione discreta numerabile di autovalori, i quali per l'appunto corrispondono ai livelli di energia dell'atomo d'idrogeno, ricostruiti a partire dai dati delle misure spettroscopiche, che forniscono le differenze di questi livelli di energia. In questo e in casi analoghi una spiegazione qualitativa dello spettro discreto è la seguente. L'attrazione dell'elettrone verso il nucleo è descritta dal potenziale $-e/r$. Si consideri un modello grossolano, nel quale il potenziale abbia la forma di una buca con le pareti molto alte. Allora la funzione ψ, che descrive lo stato dell'elettrone in questa buca, deve annullarsi fuori delle pareti e, per continuità, anche sulle pareti. La si decomponga in onde spaziali armoniche. Allora, di ogni componente, internamente alla buca di potenziale, vi dev'essere un numero intero di semionde. Ma l'energia di un'onda armonica è proporzionale al quadrato di questo numero e ha un valore definito solo nel caso armonico. Ciò spiega come in questo modello si abbia uno spettro discreto. Nel caso generale il meccanismo del discreto resta piú o meno lo stesso, benché nella dimostrazione occorra usare, non il fatto che la funzione si annulla fuori della buca di potenziale, ma solo il fatto che essa decresce all'infinito e che è a quadrato sommabile.

Numeri quantici dell'elettrone nell'atomo d'idrogeno. I livelli di energia dell'elettrone nell'atomo d'idrogeno sono enumerati in ordine crescente in modo naturale mediante il numero quantico principale $n=1, 2, 3, \ldots$ Ma ad ogni n corrisponde non una funzione ψ, ma tutto un sottospazio lineare in $\mathcal{H}$. Questo sottospazio, per la simmetria sferica dell'equazione di Schrödinger per l'atomo libero, è invariante rispetto alle rotazioni dello spazio $\mathbf{R}^3$ degli argomenti della funzione ψ. Esso si scompone in una somma diretta di sottospazi irriducibili di dimensione finita, invarianti rispetto al gruppo delle rotazioni, ognuno dei quali è caratterizzato da un numero intero l che può assumere i valori $l=0, 1, \ldots, n-1$. Questo numero quantico l può essere anche caratterizzato in termini di osservabile del momento angolare totale: sugli stati corrispondenti il suo valore è uguale a $(h/2\pi)\sqrt{l(l+1)}$. Fissando il numero quantico principale n e il numero quantico orbitale l, ancora non si ottiene in modo univoco la funzione ψ: lo spazio irriducibile corrispondente ha dimensione $2l+1$. Ponendo l'atomo in un campo magnetico diretto, per esempio, lungo uno degli assi coordinati, viene a mancare la simmetria sferica e si può studiare l'azione del gruppo delle rotazioni attorno a quest'asse sulla funzione ψ leggermente perturbata. Le rappresentazioni irriducibili del gruppo delle rotazioni sono a una dimensione e vengono numerate con numeri interi: al numero m, in presenza di una rotazione di un angolo ϑ, corrisponde la moltiplicazione della funzione ψ per $e^{im\vartheta}$. Il numero m – numero quantico magnetico – può assumere i valori $-l\leq m\leq l$. Esso corrisponde alla osservabile «componente del momento angolare nella direzione del campo», la quale può assumere i valori $hm/2\pi$. I numeri n, l, m definiscono completamente in prima approssimazione lo stato dell'elettrone.

Tuttavia in una descrizione piú esatta risulta che l'elettrone ha ancora un grado di libertà interno non classico, chiamato spin. L'osservabile dello spin ha due autovalori, $\pm 1/2$, ma essa agisce non nello spazio iniziale di Hilbert $\mathcal{H}$, ma nello spazio $\mathcal{H}\oplus\mathcal{H}$ della funzione ψ «a due componenti». In altre parole, se si vuole considerare questa osservabile, bisogna estendere lo spazio di Hilbert del sistema, aggiungendo agli argomenti della funzione ψ un nuovo argomento discreto, s, che assume due valori. Corrispondentemente bisogna ridefinire le osservabili,

in particolare l'osservabile dell'energia; a quest'ultima bisogna aggiungere un termine relativistico, la «perturbazione dello spin».

Pertanto in questa descrizione uno stato puro dell'elettrone nell'atomo d'idrogeno è caratterizzato da quattro grandezze discrete, n, l, m, s, le quali appaiono o dagli autovalori di operatori con spettro discreto, oppure quali caratteristiche di simmetria del sistema, cioè quali invarianti delle rappresentazioni irriducibili dei gruppi di simmetria.

Nubi elettroniche e principio di Pauli. La descrizione data sopra dei numeri quantici dell'elettrone si riferisce solo, a rigore, all'elettrone nel campo sferico simmetrico coulombiano del nucleo a simmetria sferica. Tuttavia, con una certa approssimazione, essa può essere usata anche nel caso di atomi con piú elettroni, ritenendo che ognuno di essi si muova in un campo che è la somma del campo del nucleo e di quello mediato degli altri elettroni. Se $\bar{x}_a$ e s_a sono le coordinate e lo spin dell'a-esimo elettrone, la funzione ψ dell'atomo dovrà essere una certa funzione $\psi(\ldots, \bar{x}_a, s_a, \ldots)$ Gli elettroni sono indistinguibili, cioè la funzione che si ottiene dalla precedente permutando gli argomenti $(\bar{x}_a, s_a)$ deve descrivere lo stesso stato. Si può dare un fondamento alla proposizione secondo la quale, per ogni permutazione degli argomenti $(\bar{x}_a, s_a)$, la funzione dev'essere moltiplicata per un fattore di fase $e^{i\vartheta}$. Dalle proprietà del gruppo simmetrico segue che questo fattore può essere uguale solo a ± 1. Il principio di Pauli afferma che esso è uguale a -1, cioè che la funzione ψ dell'atomo è antisimmetrica rispetto alla permutazione degli elettroni. Poiché essa non è nulla, in un atomo nello stato stazionario non vi possono essere due elettroni in uno stesso stato, cioè con gli stessi numeri quantici (n_a, l_a, m_a, s_a). È, questo, il famoso principio di esclusione. Le nubi ricordate nel § 1.1 sono insiemi di elettroni con numeri quantici n, l assegnati.

Spettri di massa e interazioni delle particelle elementari. Benché attualmente non esista una teoria conseguente delle particelle elementari che fornisca una spiegazione degl'invarianti discreti ad esse connessi (spettri di massa, numeri quantici), si hanno delle teorie semifenomenologiche assai soddisfacenti, nelle quali questi invarianti appaiono quali caratteristiche di rappresentazioni di diversi gruppi di simmetria.

Fotoni e quantizzazione del campo. Il campo elettromagnetico classico è dato dai valori dei vettori **E** e **H** in ogni istante del tempo e in ogni punto dello spazio. In accordo con l'idea generale della meccanica quantistica, ad essi devono corrispondere delle osservabili, cioè degli operatori dipendenti dai punti dello spazio-tempo. L'osservabile dell'energia sarà pure un operatore. Alla densità dell'energia, trasportata dal campo con frequenza ν, corrisponde, nella teoria quantistica, la densità della probabilità di osservare un quanto di energia $h\nu$, cosí che in questa interpretazione il «discreto» dei fotoni è pure legato allo spettro discreto dell'operatore dell'energia. Lo spazio di Hilbert degli stati dell'elettrodinamica quantistica, in generale, contiene dei sottospazi che corrispondono ai «diversi numeri di particelle». Essi sono sottospazi propri dell'osservabile discreta «numero di particelle». L'evoluzione del vettore di stato in questo spazio descrive i diversi processi di annichilazione e di creazione di particelle e di quanti, che avvengono nelle interazioni tra il campo e la materia. Le difficoltà matematiche della teoria dei campi quantizzati sono tutt'altro che superate. L'elettrodinamica quantistica è quella che si trova a uno stato piú perfezionato; in essa i metodi di calcolo si basano sullo sviluppo in serie di potenze nella costante di «struttura fine». Queste serie della teoria delle perturbazioni divergono, ma, limitandosi ad un numero finito di termini ed utilizzando vari metodi euristici di rinormalizzazione, si ottengono risultati che sono in ottimo accordo con gli esperimenti.

Ancor prima dell'apparizione della moderna elettrodinamica quantistica Dirac usò un'altra procedura di quantizzazione del campo per spiegare i processi di radiazione e di assorbimento dei fotoni da parte dell'atomo. Essa è basata sul fatto che un campo elettromagnetico in un dominio con pareti assolutamente riflettenti è rappresentato dalla sovrapposizione di una successione numerabile di «oscillazioni fondamentali». Le equazioni di Maxwell che descrivono questo campo coincidono formalmente con le equazioni della meccanica hamiltoniana per un sistema di infiniti oscillatori. Applicando a sottoinsiemi finiti di tali oscillatori la procedura di quantizzazione piú sopra descritta e passando al limite rispetto al numero degli oscillatori inclusi, Dirac ottenne una buona descrizione qualitativa dei salti quantici del sistema formato da atomi e campo elettromagnetico.

1.4. Difficoltà di raccordo tra teorie macroscopiche e microscopiche.

La teoria fondamentale del «mondo in grande», cioè la teoria della relatività generale, e la teoria fondamentale del «mondo in piccolo», cioè la meccanica quantistica, non concordano sia per il loro apparato formale, sia per i principî fisici sui quali queste teorie si basano. In particolare nella teoria della relatività generale non vi è posto per gli effetti quantici dell'indeterminazione e della fluttuazione probabilistica, mentre in meccanica quantistica non si considerano la curvatura dello spazio-tempo e, in generale, le interazioni gravitazionali, quest'ultime essendo, a distanze microscopiche, eccezionalmente deboli. Se si misura la forza delle interazioni fondamentali mediante la relazione tra l'energia tipica dell'interazione e la massa delle particelle interagenti, e se si assume uguale a 1 la caratteristica dell'interazione forte, si ricava che l'interazione elettromagnetica è dell'ordine di 10^{-2}, quella debole è circa 10^{-5}, mentre quella gravitazionale è dell'ordine di 10^{-38}. Tuttavia, a fondamento di tutte e due le teorie vi è la concezione generale dello spazio-tempo differenziabile, quest'ultimo intervenendo quale ambiente dell'evoluzione della metrica, della materia e dell'energia nella teoria della relatività generale, e della funzione d'onda nella meccanica quantistica. Perciò occorre prima di tutto giudicare i fondamenti fisici di questa concezione.

I punti dello spazio-tempo. Come si è già ricordato, un punto dello spazio-tempo è l'idealizzazione di un processo fisico localizzato in un piccolissimo do-

minio dello spazio e di durata brevissima. Si consideri una particella materiale di massa m. Nella teoria della relatività generale si può definire una lunghezza caratteristica, il raggio gravitazionale $a_g = 2km/c^2$, nei limiti della quale la curvatura dello spazio vicino alla particella diventa essenziale. Se si considera la particella puntiforme, il potenziale gravitazionale internamente a una sfera con tale raggio è cosí grande che da questa sfera verso l'esterno non potrebbe uscire alcun segnale; l'energia dei fotoni non è sufficientemente grande perché essi possano superare l'attrazione. Piú geometricamente, si può dire che nell'interno di questa sfera si forma un dominio dello spazio-tempo «chiuso in se stesso», cioè un «buco nero». Ciò è vero anche per una particella di raggio non nullo a, se esso è minore di a_g, ed è vero per i corpi macroscopici. Il raggio gravitazionale del Sole è circa 2,5 km, cosí che questo dominio è interno al Sole. Tuttavia buchi neri possono formarsi attorno a stelle molto dense e persino intorno ad addensamenti di materia con una densità dell'ordine di 1 g/cm^3, ma di grandi dimensioni. Nel micromondo, quanto piú esattamente vogliamo localizzare una particella, tanto minore dev'essere il suo raggio gravitazionale a_g, cioè tanto minori devono essere la sua massa m e il suo raggio a. Ma allora la densità della materia internamente alla sfera di raggio gravitazionale, per $a \sim a_g$, cresce come $1/m^2$, ciò che può superare le densità note delle particelle elementari. Si osservi che le particelle elementari con massa a riposo nulla, cioè fotoni e neutrini, non possono servire per la localizzazione spaziale, poiché si muovono con la velocità della luce. In ogni caso, nell'ambito della teoria della relatività generale la localizzazione di una particella richiede che la sua massa e il suo raggio tendano a zero.

Nell'ambito della meccanica quantistica, invece – come mostra l'analisi con l'aiuto del principio d'indeterminazione –, una particella di massa m non può essere localizzata con precisione maggiore di $h/2\pi mc$ e ciò richiede che m tenda all'infinito. Pertanto un esperimento ideale per identificare un punto dello spazio-tempo nella teoria della relatività generale e in meccanica quantistica presenta esigenze contraddittorie.

Fluttuazioni quantistiche della metrica. Se il campo gravitazionale su scala di grandezza piccola è soggetto alle ipotesi della meccanica quantistica, i coefficienti g_{ik} della metrica dello spazio-tempo devono essere trasformati nei valori di certe osservabili quantistiche e perciò diventare non completamente definiti, ma solo oscillare attorno a valori medi. La dimensione di queste oscillazioni definisce il dominio entro il quale la fluttuazione della distanza fra due punti è confrontabile con questa distanza e perciò è come se perdesse senso la localizzazione rigorosa dei punti.

È per questo che non cessano i tentativi di trovare schemi di quantizzazione, non solo dei processi fisici che si svolgono nello spazio-tempo, ma anche dello stesso spazio-tempo. Se questi tentativi saranno coronati da successo, saremo testimoni di nuovi sorprendenti fenomeni coinvolgenti il discreto/continuo nelle stesse caratteristiche fondamentali dell'esistenza.

2. *Informazione.*

2.1. Considerazioni generali.

Una parte notevole di questo paragrafo è dedicata ai sistemi viventi, da quelli cellulari e subcellulari a quelli ecologici e sociali. Per questi sistemi sono caratteristici un'eccezionale complessità di organizzazione e un alto grado di concordanza nei processi fondamentali spaziali e temporali. La coordinazione che garantisce la nascita, la crescita e lo sviluppo di tali sistemi si realizza mediante una rete complessa di flussi d'azione che verranno considerati flussi d'informazione nel senso piú esteso della parola. Il ruolo del continuo/discreto nei diversi processi di generazione, trasmissione, conservazione ed elaborazione delle informazioni verrà esposto su alcuni esempi. Saranno considerati i processi di codificazione, di trasmissione da organismo a organismo e di utilizzazione dell'informazione ereditaria che assicurano il ricambio delle generazioni biologiche e l'evoluzione del mondo vivente del nostro pianeta. Si passerà poi ai processi di regolazione nell'organismo vivente che si realizzano mediante il sistema nervoso ed endocrino. Il tema successivo sarà il «linguaggio degli animali»: i meccanismi di trasmissione delle informazioni nei sistemi ecologici. Infine si esamineranno le lingue naturali dell'uomo: scrittura e mezzi tecnici per conservare ed elaborare i dati.

Analizzando la struttura dei sistemi naturali d'informazione possiamo tentare di enunciare un problema che non si può impostare nell'ambito dei fenomeni fisici. Perché l'ereditarietà è regolata mediante un meccanismo con una componente spiccatamente discreta, quando, invece, nei processi di ricambio e di crescita il ruolo piú importante è svolto da una regolazione ormonale continua? Piú esattamente, quali sono i vantaggi e gli svantaggi dei metodi discreti e continui nella rappresentazione ed elaborazione delle informazioni con riferimento al sistema di funzioni da compiere?

Tali questioni sono state discusse attivamente negli ultimi decenni in seguito all'apparizione delle macchine calcolatrici analogiche e di quelle digitali. Nelle prime i numeri si realizzano materialmente mediante i valori di alcune grandezze fisiche quali l'angolo di rotazione di un asse, o l'intensità di una corrente elettrica, o la sua tensione, che possono variare con continuità entro certi limiti. Nelle seconde i numeri vengono rappresentati, cosí come nei testi stampati, nel sistema di numerazione binaria, o in un altro sistema, con successioni di «cifre». Le «cifre» 0 e 1 si realizzano rispettivamente con uno dei due stati stabili di un elemento elettronico. La logica dell'elaborazione delle informazioni nei dispositivi analogici in una certa misura imita i processi naturali: se l'integrale $\int_0^T v(t)dt$ significa lo spazio percorso durante il tempo T con velocità locale $v(t)$, in una macchina analogica si può porre un condensatore sotto carica con velocità $v(t)$ e poi misurare la sua carica dopo un tempo T. La logica dell'elaborazione delle informazioni nei dispositivi discreti imita i processi di calcolo dei matematici, nei quali si tratta sostanzialmente e in fin dei conti di sommare o moltiplicare cifre binarie o decimali, ripetendo questa procedura un enorme numero di volte in corrispon-

denza ad un sistema di regole complessivamente organizzato per la scelta delle operazioni successive.

Le differenze nelle caratteristiche dei calcolatori analogici e digitali sono inoltre legate alla precisione nella rappresentazione delle informazioni, alla velocità della loro elaborazione e alle caratteristiche dei dispositivi di memoria. La precisione relativa in una rappresentazione analogica può essere dell'ordine di 10^{-3} e praticamente non supera mai 10^{-5}; un aumento della precisione richiederebbe dispositivi di un'enorme complessità. La precisione relativa in una rappresentazione discreta teoricamente non ha limiti; per accrescerla si richiede solo di aumentare il numero delle cifre, oppure di realizzare metodi opportuni di programmazione. D'altro lato, per la maggior parte dei calcoli scientifici ed economici, nei dati in uscita non si richiede un'alta precisione, mentre nei dati in ingresso un'alta precisione può essere solo fittizia. L'esigenza di un'alta precisione nei risultati intermedi nasce invece dalle caratteristiche stesse dell'elaborazione discreta delle informazioni: l'enorme quantità di piccoli passi intermedi accumula gli errori. La velocità dei calcolatori digitali rispetto a quelli analogici mira anche in misura notevole a compensare la lunghezza delle procedure di calcolo. Ad ogni modo, allo stato attuale della tecnica, un bilancio generale propende decisamente per i calcolatori digitali. A ciò contribuisce non poco la loro universalità in contrapposizione alla limitatezza della classe dei problemi risolvibili da ciascun calcolatore analogico; fatto, questo, a fondamento della loro struttura. In linea di principio un grande calcolatore digitale è in grado di eseguire qualsiasi algoritmo. Infine, la realizzazione tecnica della memoria nei dispositivi discreti risolve efficacemente il problema del volume e della stabilità dell'informazione da conservare.

È interessante tentare di capire in che misura questa prevalenza dell'aspetto discreto dell'informazione nei sistemi umani di scrittura ed elaborazione dell'esperienza sia dovuta all'essenza dei fatti, o non sia piuttosto una scelta parziale delle soluzioni tecniche. John von Neumann [1958] esaminò a fondo questo problema, confrontando i principî usati nella progettazione e costruzione delle macchine calcolatrici con i principî del funzionamento del sistema nervoso centrale dell'uomo. Egli giunse alla conclusione che la logica interna del cervello dev'essere sostanzialmente diversa, e per molti aspetti, dalla logica dei nostri calcolatori. Il sistema nervoso non può conseguire un'alta precisione (superiore a 10^{-2}, o 10^{-3}) nella rappresentazione delle grandezze neppure in linea di principio. In esso non possono quindi realizzarsi quei processi profondamente conseguenti che sono caratteristici dei calcolatori digitali e, in generale, delle strutture deduttive nei metodi matematici. Nel sistema nervoso hanno un'enorme importanza le proprietà statistiche, temporali e spaziali, dei segnali e la massa dei processi paralleli informativi. La componente discreta che si osserva a livello della codificazione in impulsi-frequenze nei neuroni viene livellata attraverso una media spazio-temporale. La regolazione ormonale partecipa alla graduale e continua variazione dei parametri fondamentali che caratterizzano l'elaborazione discreta; per esempio, delle soglie di eccitazione.

Probabilmente vi sono alcune cause profonde del fatto che l'aspetto discreto

della codificazione delle informazioni prevalga nettamente a due livelli estremi: a quello piú profondo dell'ereditarietà biologica e a quello piú superficiale della comunicazione tra le persone. A quest'ultimo livello è caratteristico l'uso della lingua naturale, la quale, sul piano dell'espressione, è discreta per sua natura. L'uso della scrittura e dei sistemi speciali di segni della matematica e di altre scienze naturali sottolinea ancora di piú questa componente discreta. Nell'ambito della cultura umana è interessante mettere a confronto i linguaggi discreti della scienza con i linguaggi continui quali la pittura, la musica e la danza. La continuità del linguaggio pittorico, o musicale, sul piano dell'espressione, è correlata con l'amorficità della sua semantica (per lo meno nelle tradizioni europee moderne) e ciò in contrasto con le caratteristiche semantiche rigide del linguaggio scientifico. L'uso della lingua naturale nella letteratura artistica e, in parte, negli studi umanistici occupa una posizione intermedia tra questi due casi estremi: l'epopea di Gilgamesh o l'*Amleto* dicono qualcosa di piú che non l'insieme di parole che formano questi testi.

Il carattere discreto del linguaggio scientifico, sia sul piano dell'espressione sia su quello del contenuto, risulta indispensabile perché questo linguaggio assolva le sue funzioni sociali fondamentali, cioè affinché l'uomo possa fissare, conservare, elaborare e usare tutti quei dati sul mondo esterno che devono essere assunti in modo univoco, conservati a lungo e che devono sottoporsi ai lunghi processi dell'elaborazione deduttiva.

Ma queste stesse proprietà del linguaggio scientifico inducono ad essere cauti, soprattutto esaminando il tema di questo paragrafo: non potrebbe essere che il discreto da noi osservato nei processi naturali sia in misura notevole un artefatto da noi introdotto mediante i nostri mezzi di descrizione? Esaminando il discreto/continuo delle osservabili fisiche abbiamo visto quanto sia non-banale l'interazione tra i due aspetti e i modi come uno genera l'altro persino nei modelli matematici non molto complessi. Tra l'altro il mondo dei processi informativi è studiato assai peggio; quasi sempre siamo costretti a restare a un livello fenomenologico e troppo spesso siamo inclini a far ricorso alle analogie con i sistemi di comunicazione artificiali.

«L'informazione ereditaria è scritta in un codice lineare di venti lettere». Questa affermazione va accettata con la stessa cautela con la quale si accetta l'affermazione che la frase precedente contiene un'informazione sui successi della biologia molecolare. Si è costretti ad esprimere una verità continua mediante un aforisma discreto, ma non si deve dimenticare la differenza tra l'uno e l'altro.

2.2. Ereditarietà.

Una delle fondamentali generalizzazioni della biologia teorica è il concetto secondo il quale «tutto ciò che vive proviene dal vivente». Attualmente non si osservano autoconcepimenti spontanei della vita, benché, probabilmente, essi possano aver avuto luogo sul nostro pianeta miliardi di anni fa. Qui non si toccherà questo processo della biogenesi.

Un nuovo organismo riceve un'informazione ereditaria codificata in strutture

biologiche compatte, la quale definisce la sua formazione, il suo funzionamento, la sua crescita e, possibilmente, la sua morte, e cosí pure i meccanismi per trasmettere ai posteri l'informazione da esso ricevuta. Gli ultimi decenni hanno dato molte notizie sulla forma e il funzionamento di queste strutture, di cui si dà ora una breve esposizione.

Caratteristiche degli organismi viventi direttamente osservabili sono la crescita, il moto, l'alimentazione, le reazioni alle sollecitazioni esterne, la proliferazione. Il fondamento chimico di tutti questi eventi è il metabolismo (scambio di sostanze). Esso consiste nella trasformazione delle sostanze assimilate con l'alimentazione in strutture dell'organismo e nell'utilizzazione delle energie liberate da queste sostanze. Le reazioni delle sostanze organiche, in generale, sono particolarmente lente. Per accelerarle nella cellula viva intervengono gli enzimi. L'azione di questi ultimi in primo luogo ha un carattere catalitico: una molecola di un enzima può agire, senza modificarsi, su una quantità enorme di molecole delle sostanze reagenti (substrati). In secondo luogo gli enzimi sono altamente specializzati sia in relazione alla scelta del substrato, sia in relazione al carattere dell'azione su di esso. Il carattere dell'informazione ereditaria è definito in primo luogo dalla qualità degli enzimi sintetizzati dalla cellula e dalle reazioni che questi stessi enzimi regolano. Gli enzimi (piú esattamente, le parti delle loro molecole, cioè gli apoenzimi, responsabili del riconoscimento del substrato) sono proteine.

La molecola della proteina ha una caratteristica lineare: essa è una catena polipeptidica, i cui anelli sono le molecole di amminoacidi, piú esattamente i residui amminoacidici. La lunghezza di una tale catena può oscillare fra 100 e 30 000 anelli. Si hanno venti tipi diversi di anelli, cioè in tutto si hanno venti amminoacidi. La specifica della proteina è definita in massima parte dall'esatta successione dei residui amminoacidici nella catena polipeptidica delle sue molecole. Di regola questa successione viene letta in una direzione che nella cellula è pure univocamente definita.

Ognuno dei venti amminoacidi ha la struttura seguente. A un estremo della molecola si trova il gruppo carbossilico COOH con proprietà acide; attraverso un secondo atomo di carbonio questo gruppo è unito al gruppo amminico NH_2; inoltre a questo atomo di carbonio sono uniti un atomo d'idrogeno e il residuo amminoacidico. Il residuo, che definisce la specificità dell'amminoacido, può essere di struttura complessa, contenendo fino a venti atomi di carbonio. Quando due amminoacidi si uniscono, formando un dipeptide, si libera una molecola d'acqua e si forma un legame tra il gruppo amminico del primo acido e il gruppo carbossilico del secondo acido. In modo analogo si forma una catena polipeptidica. Una catena polipeptidica è intrecciata e ripiegata spazialmente su se stessa in modi diversi. Queste strutture si dicono secondarie, terziarie, ecc., a differenza di quella primaria che è la successione dei residui. La struttura secondaria piú importante è quella a spirale. Essa è stabilizzata mediante legami di idrogeno. I residui amminoacidici sporgono all'esterno della spirale. A una forma (α-forma) appartengono 3,7 legami peptidici per ogni spira; nell'altra forma (β-forma) vi sono 5,1 residui per ogni spira. La struttura secondaria – spirale – in certi tratti

può ripiegarsi, avvolgersi e anche ramificarsi, dando luogo alla struttura terziaria. L'unione di alcune molecole proteiche in un complesso genera la struttura quaternaria. La caratteristica di tutte queste strutture è definita dalla successione dei residui amminoacidici della molecola e cosí pure dall'ambiente nel quale si trova la molecola. A sua volta, la specificità della struttura spaziale della molecola definisce la reazione cui essa può prender parte, in particolare la scelta del substrato da parte dell'enzima.

Pertanto la molecola proteica, quale supporto dell'informazione, può essere modellata con buona approssimazione mediante una «parola» in un alfabeto di venti lettere, cioè con la successione primaria dei residui amminoacidici in questa molecola.

Tuttavia, le molecole della proteina non sono i supporti principali dell'informazione nella cellula. Fanno parte di questi supporti altri composti organici fondamentali, cioè gli acidi nucleici, detti cosí perché si trovano prevalentemente nel nucleo della cellula.

Le molecole di acidi nucleici sono pure catene lineari, i cui anelli si chiamano nucleotidi. Un nucleotide si ottiene mediante una composizione successiva di tre molecole: acido fosforico (lo stesso in tutti i casi), zucchero (una molecola di riboso, o di desossiriboso) e una base organica. In una molecola di acido nucleico vi sono solo nucleotidi con un solo tipo di zucchero; corrispondentemente gli acidi nucleici si dividono in due classi: desossiribonucleici (DNA) e ribonucleici (RNA). In qualità di basi, vi possono essere nei desossiribonucleotidi solo quattro sostanze: adenina (A), citosina (C), guanina (G) e timina (T). Anche nei ribonucleotidi possono entrare, quali basi, solo quattro sostanze: A, C, G, oppure uracile (U). Una catena polinucleotidica, cioè una molecola di acido nucleico, può contenere da alcune decine sino ad alcuni milioni di nucleotidi. Corrispondentemente una tale molecola, quale supporto dell'informazione, può essere modellata come una parola nell'alfabeto di quattro lettere: A, C, G, T per il DNA e A, C, G, U per l'RNA. Quando nella cellula si produce la sintesi di una proteina avviene un processo di ricodificazione di tali «parole nucleotidiche» in «parole amminoacidiche» delle stesse proteine. L'interpretazione di questo codice e cosí pure la spiegazione dei particolari del meccanismo di ricodificazione sono tra le scoperte fondamentali della biologia del XX secolo.

Punto di partenza sia la descrizione del codice. Per decifrare la successione dei nucleotidi bisogna suddividerla in terne, o triplette: per esempio, GAUGCU CCUUUU si trasforma nella successione di triplette (GAU) (GCU) (CCU) (UUU). A ogni tripletta di nucleotidi corrisponde un ben definito residuo amminoacidico: cosí, GAU corrisponde all'acido aspartico, GCU all'alamina, CCU alla prolina, mentre UUU corrisponde alla fenilolanina. Pertanto, GAU, GCU, CCU, UUU vengono ricodificati in asparagina, alanina, prolina e fenilolanina, precisamente in quest'ordine. Si osservi che tanto nel DNA quanto nell'RNA le triplette possibili sono di $4^3 = 64$ tipi, mentre di amminoacidi in tutto se ne hanno 20. Pertanto il codice genetico è degenere: a triplette diverse può corrispondere uno stesso residuo amminoacidico. Cosí, per esempio, a tutte le triplette GUU, GUC, GUA, GUG corrisponde uno stesso acido, la valina.

È essenziale il fatto che questo metodo di codificazione sia altamente sensibile agli errori. È vero che, sostituendo un nucleotide con un altro, cambia solo una tripletta e, corrispondentemente, un solo residuo amminoacidico. Ma, per esempio, traslando l'inizio della lettura di una sola unità, può cambiare tutta la successione di triplette e, corrispondentemente, ciò può condurre alla sintesi di una proteina assolutamente diversa. Nel nostro esempio, cominciando dalla seconda lettera, si ottiene la successione di triplette (AUG), (CUC), (CUU), (UU...) e quindi la sintesi del polipeptide metiomina-leucina-leucina-... Si produce un effetto analogo se viene a cadere un nucleotide, a cominciare dal posto dove manca, oppure se ne viene introdotto uno in piú. Questo carattere altamente discreto della scrittura, lettura, ricodificazione e reazione agli errori nel codice genetico, definisce il ruolo delle mutazioni nelle modificazioni degli organismi viventi. Le mutazioni sono variazioni casuali discontinue dell'apparato genetico che si producono per l'influenza di fattori chimici, di radiazioni, di fluttuazioni di calore, ecc. Esse possono interessare i geni e i cromosomi; consistere nella variazione, nella perdita, o nella rotazione di 180° di un tratto della catena, ecc.

Si dà ora, brevemente e con notevoli semplificazioni, la descrizione del processo stesso di elaborazione dell'informazione ereditaria.

Ogni organismo riceve dai genitori (o dal genitore) un insieme di geni. Questi sono costituiti da molecole di DNA e sono distribuiti nei cromosomi. In queste molecole di DNA è contenuta tutta l'informazione ereditaria necessaria per lo sviluppo e la vita dell'organismo. Nella forma stabile una molecola di DNA è costituita non da una, ma da due catene polinucleotidiche, che si avvolgono l'una sull'altra formando la famosa doppia elica. Questa struttura è definita in base alla legge di complementarità: i legami d'idrogeno uniscono le coppie di basi che nelle due catene si trovano una di fronte all'altra. Nel DNA sono possibili solo le coppie (AT), (CG) e quelle ad esse speculari; nell'RNA sono possibili rispettivamente (AU) e (CG). Questo principio di complementarità, che è una variante della codificazione, svolge un ruolo fondamentale in due processi: nella duplicazione e nella trasmissione dell'informazione dal DNA alla proteina.

Nel processo di divisione della cellula la spirale DNA si sviluppa e ognuna delle due catene polinucleotidiche costruisce su se stessa, come su una matrice, la catena complementare. Quale risultato si ha la duplicazione di una copia dell'informazione ereditaria.

Il processo di trasmissione dell'informazione dal DNA alla proteina si attua attraverso alcuni stadi. All'inizio il DNA (una delle sue catene) costruisce su se stesso, secondo il principio di complementarità, la catena RNA (sono possibili solo le coppie (AU), (CG), (GC), (TA)): questo RNA si chiama messaggero (m-RNA). Le triplette di nucleotidi nell'm-RNA si dicono antitriplette. In una molecola di DNA è scritta sequenzialmente l'informazione sulla sintesi di molte proteine diverse; una molecola di m-RNA è notevolmente piú corta; essa contiene l'informazione per una sola proteina, che, in base alla complementarità, si riferisce al tratto corrispondente della catena del DNA.

Una lunga molecola di m-RNA lega tra loro alcune speciali formazioni intracellulari, i ribosomi, a forma di corpuscoli tondeggianti. Essi sono le «fabbriche

delle proteine», le quali vengono sintetizzate sulle loro superfici. In particolare essi trattengono le molecole di m-RNA in uno stato disteso, ciò che rende accessibile l'informazione in esse contenuta.

Gli amminoacidi che saranno uniti nella catena polipeptidica giungono ai ribosomi uniti a molecole di RNA di tipo speciale che si chiamano RNA transfer (t-RNA). Ogni tipo di t-RNA ha in una determinata parte della molecola una tripletta speciale di nucleotidi: la sua antitripletta. L'antitripletta definisce, da un lato, quel tipo di amminoacido che questo RNA è in grado di trasportare, e dall'altro lato, quella (anti)tripletta dell'm-RNA sul ribosoma, con la quale è in grado di unirsi questa molecola di t-RNA insieme con il suo residuo amminoacidico. Dopo che a un m-RNA, in due triplette contigue, si sono unite due molecole di t-RNA con i residui, tra esse si forma un legame peptidico; una delle molecole di t-RNA si libera, si stacca dal ribosoma e può di nuovo compiere la sua funzione. In tal modo, sequenzialmente, si costruisce la catena polipeptidica. Dopo che questa si è completata, essa si divide dal ribosoma e si attorciglia, assumendo le sue strutture secondaria e terziaria.

Riassumendo, si può dire che in tutti i tratti da noi considerati, nei quali inizia la catena di trasmissione dell'informazione ereditaria, dominano le strutture discrete: le piccole unità messaggere, i nucleotidi e i residui acidi si organizzano in un sistema lineare con rigide regole di ricodificazione. Nell'ultimo stadio di questa catena il discreto si esprime mediante il principio «un gene - un enzima - una reazione». Le reazioni che avvengono nella cellula hanno una struttura spazio-temporale già cosí complessa che, per poter spiegare le loro caratteristiche continue/discrete, bisogna passare a un altro livello di descrizione.

Omettendo per necessità ulteriori particolari e tralasciando l'enorme gerarchia dei successivi livelli, non si può non accennare per sommi capi al manifestarsi di questa organizzazione discreta dell'informazione genetica a livello dell'individuo, della popolazione e della specie.

Prima di tutto la costituzione genetica che l'organismo eredita dai suoi genitori, cioè il suo genotipo, si manifesta esternamente in un insieme di caratteri visibili che costituiscono il fenotipo dell'organismo. I caratteri fenotipici possono essere morfologici (colore, grandezza, forma del corpo, o di sue parti), oppure fisiologici (presenza o assenza di un determinato enzima). Nel caso piú semplice il carattere fenotipico è controllato (regolato) da un solo gene di un dato carattere. Nella riproduzione bisessuale l'organismo riceve dai genitori due geni di ciascun sesso: uno paterno e l'altro materno. Essi possono corrispondere a forme diverse del carattere fenotipico. Per esempio, nel genotipo di uno stesso uomo vi può essere il gene degli occhi castani, ricevuto dal padre, e il gene degli occhi celesti, ricevuto dalla madre. Uno di questi geni può definire completamente il carattere e, in tal caso, esso si dice dominante, mentre l'altro si dice recessivo. Può anche accadere che l'azione contemporanea di due geni conduca a una scala di gradazioni del carattere fenotipico. È, questo, uno dei fattori della continua variabilità del fenotipo. Un altro fattore è l'influenza dell'ambiente esterno. Nei mammiferi uno dei meccanismi genetici dove l'azione discreta si manifesta in modo piú evidente è quello della regolazione del sesso.

L'evoluzione di questa o quella forma degli organismi viventi avviene nel processo di avvicendamento delle nuove generazioni d'individui, viventi in un determinato territorio e che s'incrociano tra loro: si tratta di una popolazione genetica. Il genotipo di ogni individuo è unico, ma la frequenza media di varietà (allele) di ogni gene in una grande popolazione è relativamente stabile. Per l'evoluzione è essenziale alterare questa stabilità e ciò avviene grazie ai processi di mutazione, alle rigenerazioni selettive e, nelle piccole popolazioni, alle fluttuazioni casuali del fondo genico. Infine, può verificarsi un afflusso di geni da un'altra popolazione. La media delle variazioni discrete del fondo genico, rispetto a una grande popolazione, produce l'effetto di un'evoluzione continua. Una brusca variazione delle condizioni esterne che conduca a una selezione e a una successiva rigenerazione d'individui, con uno specifico genotipo (apparso, probabilmente, quale risultato di mutazioni), si presenta come un salto evolutivo. I salti evolutivi si notano in modo particolare negli organismi primitivi con un fondo genico povero. I virus, costituiti quasi esclusivamente da materiale genetico, subiscono in modo particolarmente facile variazioni e salti. Sono largamente noti gli esempi forniti dai virus delle infezioni influenzali.

Esiste l'ipotesi che gli uccelli, considerati quale classe zoologica, si poterono formare da antenati pennuti a conclusione di un'unica mutazione, nella quale si è accorciata la loro lunga coda (simile a quella dei rettili con le penne), ciò che migliorò nettamente le loro caratteristiche aerodinamiche. La possibilità di una tale mutazione si dimostra con l'esempio dei gatti senza coda dell'isola di Man.

Un gruppo di popolazioni con un fondo genico cosí comune che tra loro siano possibili gl'incroci si chiama specie. Le frontiere tra le specie, pur non essendo barriere assolute, sono definite in modo sufficientemente netto dall'incompatibilità dei genotipi, i quali ostacolano la riproduzione di forme intermedie. La presenza di queste frontiere è ancora una manifestazione del discreto nell'informazione genetica. Nella formazione delle specie possono influire diversi fattori isolanti, che ostacolano l'incrocio tra sottogruppi della popolazione e che, nel processo dello sviluppo evolutivo, favoriscono la divergenza dei loro fondi genici in direzioni diverse.

2.3. Il sistema nervoso e il sistema endocrino.

La trasmissione delle informazioni all'interno di un organismo pluricellulare complesso viene attuata fondamentalmente da due sistemi specializzati: quello nervoso e quello endocrino. Il sistema nervoso controlla prevalentemente le reazioni veloci (la cui durata non supera il millisecondo), mentre quello endocrino controlla le reazioni lente, di lunga durata. In entrambi i sistemi l'interazione intercellulare si attua mediante la secrezione di sostanze speciali, cioè è quasi di tipo analogico. Nelle cellule nervose la trasmissione intracellulare delle informazioni avviene mediante impulsi elettrici, aventi una caratteristica discreta. Tuttavia la stessa codificazione delle informazioni usa proprietà statistiche spazio-temporali degli impulsi.

Il sistema nervoso. Negli organismi pluricellulari esso consiste di cellule specializzate: i neuroni. In un tipico neurone si individuano le parti seguenti: *a*) un'appendice recettiva (dendriti); *b*) un luogo per la generazione degli impulsi nervosi (solitamente il tratto iniziale di una lunga appendice, l'assone); *c*) un elemento conduttore, lungo il quale scorrono velocemente gl'impulsi, l'assone; *d*) una parte terminale dell'assone, dove avviene la secrezione dei mediatori verso l'appendice recettiva di un altro neurone; *e*) il corpo del neurone (pericario, soma), che sostanzialmente svolge funzioni trofiche. L'unità funzionale del sistema nervoso viene assicurata da legami interneuronici, realizzati mediante sinapsi. Nelle sinapsi tra i neuroni vi è una sottile fessura fra il tratto di membrana della parte terminale del neurone presinaptico e il tratto di membrana della parte recettiva del neurone postsinaptico. Il numero di neuroni in tutto il sistema nervoso dell'uomo è dell'ordine di 10^{10}. I neuroni della spina dorsale e del cervello formano il sistema nervoso centrale; gli altri formano il sistema nervoso periferico. Il percorso tipico di un segnale dall'ambiente esterno è il seguente: le cellule recettive (recettori cutanei, cellule del gusto, recettori della vista, ecc.), sottoposte all'azione di uno stimolo esterno, inviano attraverso le sinapsi un segnale ai dendriti terminali dei neuroni, i quali portano il segnale al sistema nervoso centrale. In questo percorso i neuroni si dicono afferenti. Nella direzione inversa i neuroni efferenti trasmettono i segnali dal sistema nervoso centrale agli organi esecutivi: muscoli, ghiandole, ecc. Nello stesso sistema nervoso centrale avvengono i processi fondamentali di ricezione, coordinazione, elaborazione e conservazione dell'informazione.

Considerando piú da vicino il modo di rappresentare e trasmettere le informazioni da parte di un neurone, si rileva che il fondamento fisico è costituito dagl'impusi nervosi, o potenziali d'azione, che sono variazioni di breve durata del potenziale elettrico tra gli elettroliti da ambo i lati della membrana neuronica. Queste variazioni si propagano attraverso le appendici del neurone e hanno forma d'impulsi. Secondo le moderne interpretazioni, gl'impulsi che corrono lungo le fibre nervose sono omogenei, nel senso che la loro forma e grandezza non contiene informazione (almeno in determinati limiti); essenziale, invece, è la frequenza degl'impulsi e cosí pure l'insieme delle fibre nervose attive. Pertanto, di per sé gl'impulsi hanno un carattere discreto, mentre la loro densità spaziale e temporale è relativamente continua, ciò che assicura, per esempio, la possibilità di una sottile graduazione delle reazioni motorie. Una secoda componente essenzialmente continua del processo di elaborazione delle informazioni viene realizzata nelle sinapsi. Un impulso nervoso, quando giunge alla terminazione presinaptica, provoca la secrezione di sostanze speciali, i mediatori. Il mediatore si diffonde attraverso una stretta fessura sinaptica e agisce sulla membrana del neurone contiguo postsinaptico. Quest'azione può essere di carattere eccitante o inibitorio, provocando un aumento o una diminuzione del potenziale della membrana postsinaptica. Le sinapsi inibitorie sono di eccezionale importanza. In certe intossicazioni, quando la loro azione viene bloccata, il sistema nervoso viene preso da forti scariche convulsive in grado di uccidere l'organi-

smo. Naturalmente ciò è solo la manifestazione esterna di alterazioni assai gravi nei processi fini di elaborazione delle informazioni, dove inibizione e «censura» sono non meno importanti che la produzione di segnali. Nel sistema nervoso centrale dei mammiferi giungono ad ogni cellula postsinaptica le terminazioni presinaptiche di alcuni neuroni: parte sono inibitori e parte sono eccitanti. Le sorgenti dei segnali originali possono trovarsi in parti molto lontane dal sistema nervoso.

I segnali che giungono a un neurone vengono presi in conto con pesi diversi; il segno di questi è determinato dal carattere della sinapsi (eccitante, o inibitoria), mentre la grandezza assoluta dipende dalla distanza della sinapsi dal luogo di generazione degl'impulsi. Soprattutto le azioni eccitanti e inibitorie hanno un certo tempo di attenuazione. Se la somma delle eccitazioni supera una certa soglia, il neurone comincia esso stesso a generare gl'impulsi. In conclusione ogni neurone agisce come un sommatore non-lineare e un trasformatore di frequenze.

Tutti questi meccanismi dànno luogo alla possibilità di tipi diversi di elaborazione delle informazioni, realizzate in particolare nei nuclei sensoriali di commutazione. L'informazione può subire commutazioni in un verso o in un altro; può essere intensificata, indebolita o filtrata, di modo che alcune caratteristiche risultino attenuate e altre intensificate; può essere inviata alla memoria, oppure a sollecitare una reazione immediata; infine può essere ricodificata, cioè trascritta su un altro supporto materiale. I particolari di questi processi sono incredibilmente vari e variabili; la non-linearità è un loro tratto caratteristico; l'abbondanza dei modelli matematici può essere confrontata solo con la loro inadeguatezza. Molte generalizzazioni sono suggestive per la loro attraente semplicità; è di Mountcastle l'audace aforisma «Il cervello è un operatore lineare», con ciò supponendo che tutta la non-linearità resti a carico della periferia. Ma se ne sa troppo poco.

Si riportano, quale esempio, alcuni particolari sulla codificazione iniziale della sensazione di caldo-freddo da parte dei termorecettori cutanei. Qui si hanno due concezioni contrapposte sulla presenza di recettori di uno oppure di due tipi. La dipendenza della frequenza degl'impulsi dalla temperatura, dopo un periodo di adattamento, ha la forma di una curva con due massimi, ciò che, probabilmente, testimonia a favore dell'ipotesi di recettori di due tipi. Vi sono poi altri argomenti. D'accordo con questa teoria e secondo le misurazioni, i recettori del freddo, dopo l'adattamento, rispondono con una frequenza fino a 10 impulsi/secondo nel tratto fino a 43 °C con un massimo intorno a 28 °C; i recettori del caldo rispondono con una frequenza fino a 35-40 impulsi/secondo nel tratto da 30 a 50 °C, con un massimo intorno ai 45 °C. Ogni frequenza degl'impulsi segnala due possibili temperature solo per un tipo di recettori; per esempio, il recettore del freddo con 7 impulsi/secondo segnala 15 °C e 35 °C, però a 35 °C s'inseriscono anche i recettori del caldo, ciò che permette di distinguere queste due temperature.

Se si considerano i processi di transizione, si manifestano nuovi fenomeni. Il raffreddamento dei recettori del freddo da 20,5 a 15,2 °C, oppure da 35 a 31,5 °C

provoca una breve, ma notevole accelerazione degl'impulsi; avviene qualcosa come la «differenziazione» della temperatura rispetto al tempo nel significato matematico della parola. Tuttavia, la successiva stabilizzazione della frequenza degl'impulsi, in questi intervalli, avviene con modalità diverse; essa è piú bassa con il decrescere della temperatura nel dominio «caldo» ed è piú alta con il crescere della temperatura nel dominio «freddo».

Infine, le caratteristiche dei termorecettori hanno rivelato proprietà di cosiddetta isteresi; le misurazioni della frequenza degl'impulsi durante il raffreddamento della cute da 38 °C a 14 °C e il successivo riscaldamento fino alla stessa temperatura iniziale hanno dato curve diverse sul tratto ascendente e discendente, benché raffreddamento e riscaldamento siano stati effettuati gradatamente, con intervalli per l'adattamento.

Il sistema endocrino. Le variazioni prolungate di adattamento (legate al metabolismo, alla crescita, alla riproduzione), le concentrazioni di diverse sostanze nel sangue e dei liquidi esocellulari sono regolate da sostanze speciali: gli ormoni, distribuiti nell'organismo dal flusso del sangue. Alcuni ormoni sono prodotti da neuroni (le cosiddette cellule neurosecernenti). Gli ormoni sono molto attivi e la loro concentrazione nel sangue non è grande. Gli ormoni sono prodotti dalle ghiandole endocrine, quali la tiroide, l'ipofisi, le ghiandole surrenali. Il pancreas, le ovaie e i testicoli producono, oltre agli ormoni, anche secrezioni immesse attraverso i cosiddetti condotti escretori.

Alcuni ormoni influiscono sul metabolismo in tutte le parti del corpo. Tale è la tirosina, prodotta dalla tiroide, la quale accresce l'intensità del metabolismo fondamentale. La parte anteriore dell'ipofisi produce l'ormone della crescita, che agisce sul metabolismo proteico, lipidico e dei carboidrati e che regola la crescita delle ossa e quella generale del corpo.

Tuttavia, la maggioranza degli ormoni ha specifici «destinatari», benché il flusso del sangue li distribuisca per tutto l'organismo. Cosí la mucosa dell'intestino duodenale produce una secrezione che agisce solo sul pancreas, stimolando la produzione di succo pancreatico. Le cellule che ricoprono i follicoli dell'ovaia femminile producono l'estradiolo, il quale stimola lo sviluppo e la conservazione dei caratteri sessuali femminili.

Nei casi piú elementari l'ormone può essere considerato quale semplice supporto dell'informazione, o dell'istruzione, scritta in forma «analogica». Tale è in prima approssimazione l'azione della secretina.

Tuttavia, nella maggior parte dei casi le azioni ormonali sono riunite in un sistema con cicli di retroazioni e, qualche volta, connesse evidentemente con il sistema nervoso.

Per esempio, la concentrazione di glucosio nel sangue viene regolata da due ormoni: l'insulina e il glucagone. La concentrazione di glucosio nel sangue cresce dopo un'alimentazione ricca di carboidrati. Una parte del glucosio entra nel fegato, dove viene trasformato in riserva di glicogeno. Un aumento del livello di glucosio provoca la secrezione dell'insulina da parte di speciali β-cellule delle isole del pancreas. L'insulina accresce rapidamente il trasporto di glucosio nelle

cellule dei muscoli scheletrici e dei tessuti adiposi. Se la concentrazione del glucosio nel sangue scende sotto il livello ottimale, comincia la secrezione del glucagone da parte delle cellule del pancreas. Il glucagone attivizza un sistema speciale del fegato, nel quale il glicogeno di riserva viene trasformato, in alcune tappe, in glucosio e immesso nel sangue.

Ancora piú sottilmente viene bilanciato il meccanismo di regolazione dei cicli mestruali in alcuni mammiferi. Il concepimento è possibile solo quando un determinato tipo di comportamento sessuale esterno della femmina concorda nel tempo con l'ovulazione e lo stato della mucosa del suo utero. Alla regolazione di questi eventi partecipano almeno otto ormoni e il sistema nervoso centrale.

Notiamo infine l'esistenza di mezzi chimici per trasmettere l'informazione da un organismo a un altro. Si chiamano ferormoni una classe generale di sostanze che si diffondono nell'ambiente esterno da ghiandole endocrine e, recepite attraverso gli organi di olfatto e di gusto, influiscono sul comportamento e sulle funzioni fisiologiche di altri individui, generalmente della stessa specie. Fanno parte dei ferormoni gli attrattivi sessuali (nelle farfalle), i marcatori di tracce e i segnali di allarme (nelle formiche). Queste sostanze influiscono immediatamente sul comportamento degli individui. La regolazione della crescita delle cavallette e la regolazione del numero deg'individui sessuati e dei soldati nelle colonie di termiti si realizzano anch'esse mediante i ferormoni, l'azione dei quali è piú prolungata nel tempo e dà inizio a un'intera catena di fenomeni fisiologici complessi.

I ferormoni, similmente agli ormoni, agendo in concentrazioni molto piccole, compiono precise funzioni segnaletiche. Cosí, la femmina del baco da seta produce il suo attrattivo sessuale, il bombicolo, e lo emette nell'aria. Con un vento leggero le molecole di bombicolo di una sola femmina possono diffondersi in un'area dell'ordine di un chilometro quadrato. Queste molecole agiscono su recettori disposti sulle antenne del maschio. Esperimenti e semplici calcoli mostrano che l'azione di 10^2-10^3 molecole di bombicolo è già sopra la soglia d'eccitazione.

Anche negli animali superiori hanno una determinata importanza i segnali chimici con modulazione continua, ma con azione relativamente discreta. Tuttavia in essi cominciano a dominare i segnali recepibili dalla vista e dall'udito che si passa ora ad esaminare.

2.4. Sistemi di segnalazione negli animali.

Tutta l'informazione recepibile dagli organi di senso si può suddividere, grosso modo, in due classi: informazione dall'ambiente esterno e informazione dall'interno. Qui verrà considerata solo la prima classe; in essa si distingue una particolare sottoclasse di segnali, quelli recepibili dagl'individui della stessa specie. Le componenti discrete sul piano dell'espressione (segnali) e la semantica (reazioni comportamentali) formano il linguaggio, che è specifico per una determinata specie nella sua interazione con il mondo esterno e nel comportamento sociale.

Per noi è interessante appunto la possibilità d'individuare queste componenti discrete sullo sfondo di un'enorme variabilità di tutte le possibili azioni.

Moltissimi fatti sperimentali e generalizzazioni teoriche sono stati ottenuti in questo campo dalla scuola etologica di Lorenz. Secondo Lorenz, le forme stereotipe, istintive del comportamento, condizionate ereditariamente, formano i componenti elementari di quelle azioni integrali quali il procacciamento del cibo e la riproduzione. Per esempio, il procacciamento del cibo da parte del lupo si compone di una catena di azioni istintive: fiutare, braccare, inseguire, uccidere la vittima.

In generale un'azione istintiva è la risposta a uno specifico segnale esterno. Tuttavia la soglia dell'eccitazione varia a seconda dello stato dell'animale. Una gatta in un luogo con molti topi inizialmente li prende per mangiare. Ma, dopo essersi saziata, essa può continuare a prenderli senza ucciderli e poi scompare persino l'atto della cattura; la gatta si avvicina silenziosa solo ai topi piú lontani, senza far caso a quelli piú vicini. Al contrario, l'assenza di eccitanti, che innestino un'azione istintiva, può abbassare la soglia sino a zero e condurre l'animale allo stato di ricerca (*Appetenzverhaltung*). È proprio questo lo stato favorevole per l'ammaestramento e, in particolare, per estendere il «linguaggio» comprensibile all'animale.

È caratteristico il fatto che l'innesto di un'azione istintiva sia provocato dalla corrispondente situazione non in modo generico, ma da poche speciali caratteristiche elementari di questa situazione, dette stimoli.

Gli uccellini fuggono alla vista di animali con occhi grandi. In una serie di farfalle, nel processo della loro evoluzione, ciò produsse l'apparizione di dispositivi di difesa sotto forma di macchie a forma di grossi occhi, che diventano visibili quando le ali sono aperte.

A primavera, dopo la scelta del territorio, il maschio dello spinarolo attacca un altro maschio che si avvicini. Lo stimolo è dato da una macchia rossa sul ventre di quest'ultimo e l'attacco può essere suscitato da un qualsiasi oggetto con sotto una macchia rossa.

Questo carattere discreto degli stimoli si manifesta in modo particolarmente chiaro con l'esistenza di «ultrastimoli» che suscitano reazioni molto piú intense che in condizioni normali. Un uccello può smettere di covare le proprie uova a favore di altre piú grosse, di specie diversa, o persino a favore di pietre a forma d'uovo, cosí grosse che è impossibile covarle. Il becco aperto del grosso cuculo appena nato è un ultrastimolo che induce gli uccelli canterini ad alimentarlo.

Gli stimoli di comunicazione all'interno di una specie, che si evolvono insieme agl'istinti che essi stessi sollecitano, possono assumere forme assai specifiche e stupefacenti. Cosí il corteggiamento e l'accoppiamento dei pesci e degli uccelli è accompagnato da loro vivaci colorazioni e movimenti. Questi stimoli si chiamano *Auslöser*. La loro apparizione ed evoluzione possono essere chiarite solo dal loro carattere discreto e segnaletico.

I linguaggi di comunicazione interni a una specie possono contenere anche una componente attiva, costituita da un sistema di segnali visivi e sonori, emessi dall'individuo in determinate situazioni.

Un noto esempio è il linguaggio delle api, studiato da Frisch, con il quale esse si trasmettono l'informazione del reperimento di una fonte di cibo (cfr. fig. 1). L'ape operaia (bottinatrice), che ha trovato il cibo a una distanza di cinquanta-cento metri dall'arnia, compie al ritorno una «danza circolare». Essa è il segnale alle altre api per una ricerca a caso in un intorno vicino all'arnia. Se la distanza dal cibo è superiore, fino ad alcuni chilometri, l'ape esegue la cosiddetta «danza oscillante», con la quale vengono indicate due caratteristiche discrete della sorgente di cibo – la distanza e la direzione – e ognuna di esse viene descritta quantitativamente. L'ape descrive una curva all'incirca a forma di otto con un tratto centrale molto largo percorrendo il quale essa «vibra» il corpo la-

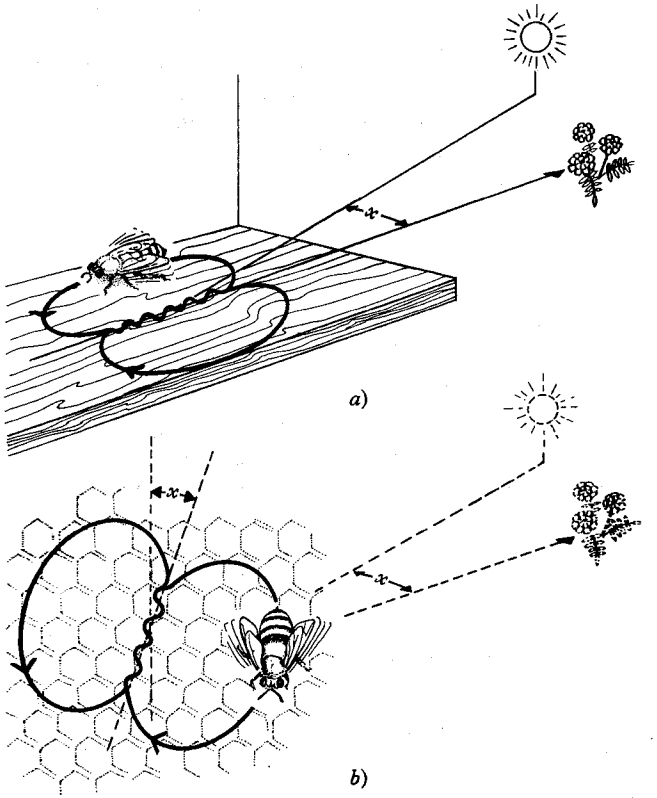


Figura 1.

Danza oscillante dell'ape domestica. La corsa indica un punto di rifornimento posto venti gradi a destra del Sole quando l'ape lascia il nido. Se l'ape effettua la danza fuori dell'alveare (*a*), il tratto rettilineo della danza indica direttamente la sorgente nutritiva; se compie la danza entro l'alveare (*b*), si orienta per gravità e il punto direttamente sopra la testa raffigura la perpendicolare. L'angolo x (venti gradi) è identico per le due danze.

teralmente, con lo spostamento massimo alla punta dell'addome e il minimo nel capo. La distanza della fonte alimentare viene codificata dal numero di passaggi dell'ape lungo il tratto centrale e dal numero di vibrazioni dell'addome per ogni secondo; con il crescere della distanza diminuisce il numero dei passaggi, ma aumenta la frequenza delle vibrazioni. La direzione della fonte alimentare viene codificata dalla direzione del passaggio lungo il tratto centrale rispetto alla verticale del luogo; un passaggio parallelo alla verticale verso l'alto corrisponde a un volo verso il sole; verso il basso, nella direzione contraria; gli angoli intermedi corrispondono a direzioni intermedie.

Il comportamento nuziale dei pesci e degli uccelli viene regolato da segnali visivi, dal linguaggio delle pose e dei loro avvicendamenti, organizzati in particolari danze tipiche della specie.

I Primati usano sistemi di segnali acustici assai complessi, comprendenti fino ad alcune decine di unità «concettuali». Per esempio, nel «linguaggio» delle bertucce sudafricane sono stati individuati segnali acustici indicanti minaccia aggressiva (abbaiamento), vicinanza di estranei, non-aggressività (sbaciucchiamento), energico avvertimento per un grosso predatore, ecc. Funzioni analoghe sono svolte da codici e azioni segnaletiche quali la posizione delle estremità, della coda, delle orecchie. Gli specialisti sottolineano che l'informazione trasmessa con questi segnali si riferisce in primo luogo a una disposizione emozionale di chi trasmette in quel momento, mentre la ricezione del segnale influisce prima di tutto sulla disposizione emozionale di chi lo riceve. Perciò la semantica di questi segnali, appena descritti, è la «semantica per un osservatore esterno». Ciò nonostante la sua realtà è incontestabile.

La struttura a segni dei sistemi biologici di comunicazione descritti induce a postulare l'esistenza, nel sistema nervoso, di meccanismi specifici, responsabili del riconoscimento e dell'elaborazione di questi segni. Probabilmente le possibilità di questi stessi meccanismi permisero e assicurarono all'uomo l'uso della parola e della scrittura. La specificità di questi sistemi è condizionata in primo luogo da un'incredibile estensione dell'universo concettuale dell'uomo, rispetto alla limitatezza del numero di segnali elementari tra loro distinguibili con sicurezza. Perciò le unità sul piano del contenuto devono essere trasmesse mediante combinazioni di unità sul piano dell'espressione. Queste combinazioni devono essere normalizzate e semanticamente orientate sul mondo esterno e non su una semplice trasmissione dello stato interno; esse devono essere sufficientemente discretizzate, affinché siano percepite con precisione sullo sfondo dei rumori e delle varianti individuali. Tutto ciò caratterizza il linguaggio umano.

2.5. Linguaggio parlato, linguaggio scritto e matematica.

Il linguaggio parlato a diversi livelli viene descritto con vari sistemi di unità discrete: fonemi, sillabe, parole, proposizioni. Il fonema, dal punto di vista acustico, è l'intervallo sonoro minimo univocamente identificabile nei limiti della sua mutabilità; quest'unità, nelle moderne trascrizioni fonetiche, viene indicata con un solo segno. Il fonema, dal punto di vista del sistema del linguaggio, è

l'unità minima in grado di compiere la funzione di distinzione del significato. Gli stessi fonemi possono essere distinti mediante caratteri differenziali ancor piú elementari, quali vocalità-consonanza, afonicità-sonorità, ecc., una parte dei quali sono rilevanti in alcune lingue e non lo sono in altre. La descrizione moderna dei fonemi è il risultato di una lunga e difficile analisi, che ha permesso di distinguere caratteristiche astratte e discrete nel flusso del discorso, questo essendo acusticamente molto complesso e avendo inoltre un dominio di possibile variabilità assai continuo, ciò che si rende evidente quando si ascolta una lingua straniera, o quando si compiono indagini elettroacustiche.

Le lingue non scritte delle società a struttura tribale erano caratterizzate da una grande mutabilità a tutti i livelli, sia nello spazio che nel tempo. La componente sonora del discorso e il vocabolario si rinnovavano velocemente a causa della mancanza di un consolidamento scritto e per l'azione di diversi tabú e dell'isolamento spaziale. Il frazionamento dialettale assunse la forma di una famiglia diffusa di lingue. Mikluho Maklai, che studiò le lingue papua nella seconda metà del XIX secolo, notò che in villaggi a quindici minuti di cammino il vocabolario poteva già differire in alcune decine di parole. Gli abitanti di villaggi a un'ora di cammino l'uno dall'altro a volte quasi non si capivano tra loro, mentre per un viaggio di oltre un giorno potevano occorrere due-tre interpreti. Anche attualmente si nota un'analoga «continuità e mutabilità linguistica». I popoli dell'Africa a sud del Sahara usano circa duemila lingue e dialetti diversi.

Nel consolidamento e nella normalizzazione delle caratteristiche discrete delle lingue ha avuto un'enorme importanza la scrittura, nelle sue forme sillabiche e alfabetiche. Tuttavia è essenziale osservare che i sistemi piú antichi, pittografici e ideografici, non fissarono il discorso sonoro, ma direttamente la semantica della comunicazione; essi furono «una seconda lingua» dell'uomo. Perciò, considerarli parallelamente alla scrittura, a rigore, non è corretto. Le unità di espressione in queste scritture corrispondono direttamente alle unità di contenuto, tralasciando l'aspetto fonetico. La sintassi dei sistemi pittografici è prossima alla «sintassi del significato» e non alla «sintassi del discorso»; in particolare essa può essere assai lontana dalla linearità.

Un tale modo di trascrivere i significati ha alcuni vantaggi essenziali, particolarmente quando è applicato a sfere semantiche specializzate. Il discorso sonoro acquista in universalità, perdendo qualche volta nella comodità dell'espressione. Nei moderni sistemi di notazioni delle scienze noi osserviamo componenti ideografiche essenziali; tale è il linguaggio internazionale delle formule chimiche, oppure, in un campo molto strettamente specializzato, il linguaggio dei «diagrammi di Feynmann» in fisica. Questi sistemi realizzano una trasmissione diretta dell'informazione al cervello, la cui espressione a parole potrebbe essere estremamente difficoltosa. Da questo punto di vista è interessante mettere a confronto il linguaggio delle formule di struttura con il linguaggio parallelo, a parole, della nomenclatura chimica. La ricodificazione di una denominazione lineare quale «3-(-dietilammino-etil)-4-metil-7-carbetossi-metossi-2-osso-(1,2-cromo) idrocloruro» nella formula di struttura e viceversa richiede l'impiego di algoritmi complessi. Di piú, l'importanza di tali sistemi ideografici non si esaurisce assolutamente nella comodità di fissare l'informazione, ma anche perché si accresce notevolmente la possibilità di rielaborarla. Si toccherà piú avanti quest'aspetto, esaminando il linguaggio simbolico della matematica.

La scrittura geroglifica è un anello di passaggio tra quella ideografica e quella fonetica. Nelle sue forme piú sviluppate, alle unità elementari dell'espressione corrispondono parole del discorso parlato, oppure nozioni definite cosí esattamente che la loro trasmissione nel discorso è limitata dalla scelta di alcuni sinonimi, mentre la sintassi dell'espressione grafica è vicina a quella del discorso ed inoltre è lineare. Ma il numero delle unità espressive è estremamente grande.

La scrittura sillabica e poi alfabetica nasce da quella geroglifica quando il centro di gravità si sposta dal significato del segno al suo suono nella lingua data. La scrittura assume sempre piú un carattere enigmatico a mano a mano che la sintassi, a livello delle parole, viene sostituita dalla sintassi a livello delle sillabe e dei fonemi, staccati dai loro significati primitivi e incorporati in un sistema parallelo al sistema del discorso sonoro.

Insieme e contemporaneamente si rafforza l'influenza inversa del linguaggio scritto sul linguaggio parlato. La grammatica sanscrita, elaborata in India oltre duemila anni fa, rigorosa e molto vicina a quelle moderne in base a principî scientifici, aveva già un aspetto normativo eccezionalmente importante, che mirava a conservare e trasmettere con precisione da una generazione all'altra i testi sacri e che agevolò la conservazione di questa lingua fino ai nostri giorni.

E cosí la scrittura fissa il sistema della lingua, sottolineando i suoi aspetti discreti, e fissa i testi socialmente importanti in questo sistema, svolgendo il ruolo di una memoria stabile esterna dell'umanità. Ma la scrittura fa di piú. La logica della lingua, fissata al di fuori della consapevolezza del portatore di questa lingua, apre nuove possibilità di generare testi e idee. La moderna sociolinguistica, che risale a Humboldt, Sapir, Whorf e altri ricercatori, studia in che misura la lingua definisce il comportamento sociale delle persone e le loro forme di conoscenza del mondo esterno. Senza soffermarsi su questo, si accennerà al ruolo del linguaggio della matematica nello sviluppo della matematica stessa e della fisica.

Il linguaggio delle formule realizza l'esteriorizzazione del pensiero, per cui frammenti notevoli del processo mentale possono essere prodotti (e non soltanto riprodotti) su supporti esterni dell'informazione, che vanno da un pezzo di carta a una lavagna con il gesso, fino ai calcolatori elettronici. E, nei calcoli manuali, la scrittura delle formule non compie solo una funzione simile al ruolo della memoria e dell'unità aritmetica dei calcolatori, ma anche quella di unità di controllo (non-deterministica) nei confronti del cervello del matematico. L'aspetto delle formule, la loro struttura e l'analogia con altre formule suggeriscono al matematico e al fisico la scelta dell'ulteriore direzione dell'indagine e la scelta della nuova semantica per il linguaggio matematico «semanticamente vuoto», o, piuttosto, «semanticamente polimorfo». La famosa storia dell'«equazione d'onda» fornisce di questo fatto illustrazioni stupefacenti. Essa fu ricavata inizialmente per descrivere le oscillazioni di una corda, o di una membrana. Maxwell, che la ricavò mediante calcoli formali dalle sue equazioni iniziali che collegano le varia-

zioni dei campi elettrico e magnetico, dedusse la possibilità della propagazione delle onde elettromagnetiche. Note costanti fisiche, presenti nelle equazioni di Maxwell, permisero di calcolare la velocità di propagazione di queste onde. Poiché essa risultò uguale alla velocità della luce, Maxwell suggerí l'audace congettura che la luce visibile avesse la stessa natura delle onde elettromagnetiche. E le stesse onde e la natura ondulatoria della luce furono poi scoperte sperimentalmente. Schrödinger scrisse la sua equazione d'onda inizialmente nella ricerca di un comodo schema matematico per chiarire gli effetti quantistici discreti. Il significato della funzione d'onda, che in quest'equazione appare quale incognita, inizialmente non era assolutamente chiaro. La sua successiva interpretazione quale «onda di probabilità», o, piú esattamente, onda di ampiezza (complessa) di probabilità, condusse a introdurre nell'universo dei concetti una struttura del tutto nuova e straordinaria, che ha cambiato sia la concezione del mondo che il modo di vivere dell'uomo moderno.

Non è possibile immaginare l'avvento di queste scoperte senza i sistemi simbolici discreti della scienza.

3. *Modelli matematici complementari.*

Nel § 1 è stata descritta brevemente una serie di modelli matematici, nei quali si considerano le interazioni tra il continuo e il discreto in molteplici situazioni fisiche. Oltre ai due casi «puri», tipo varietà differenziale da un lato e insiemi con topologia discreta (solitamente finiti, o numerabili) dall'altro, apparivano strutture piú complesse, dove gl'invarianti discreti si manifestavano come provenienti da strutture continue, ad esempio lo spettro di un operatore lineare, oppure le caratteristiche topologiche del quadro delle fasi di un sistema dinamico. In questo paragrafo, senza pretendere la completezza, saranno considerate ancora alcune strutture matematiche che, fino ad ora, hanno trovato relativamente poche, o addirittura nessuna applicazione nelle teorie fisiche, ma che, è probabile, hanno interesse per il futuro.

Anzitutto vanno sottolineati due aspetti.

In primo luogo, nell'ambito della matematica teorico-insiemistica, l'aspetto discreto è quello primario; ogni insieme è la collezione dei suoi elementi in nessun modo legati tra loro, ad eccezione, eventualmente, della relazione d'inclusione – nell'universo di Zermelo-Fraenkel. L'aspetto della continuità può essere introdotto in un insieme quale struttura supplementare, individuando una classe di insiemi speciali, gli *aperti*, che definiscono lo spazio topologico. Si potrebbero immaginare i fondamenti della matematica con un'altra assiomatica, nella quale l'aspetto continuo fosse posto fin dall'inizio. Quale tentativo estremamente ingenuo si potrebbe, ad esempio, assiomatizzare la relazione «essere parte» invece di «essere elemento». Tuttavia all'autore non sono noti seri lavori in questa direzione. L'approccio categoriale ai fondamenti della matematica porta in primo piano, evidentemente, le rappresentazioni non connesse direttamente con la continuità.

In secondo luogo tutti i modelli della continuità usati funzionano sostanzialmente con insiemi infiniti, benché la definizione di spazio topologico non richieda che esso sia infinito, ed esistano, infatti, esempi importanti di spazi topologici finiti, persino con soli due punti (uno chiuso, l'altro aperto). In fisica appaiono solitamente strutture «quasi continue» su insiemi finiti nell'ipotesi che questi insiemi siano cosí grandi da poter essere riguardati come se fossero «quasi infiniti». In realtà le risorse matematiche delle teorie moderne sono piú estese.

3.1. Topologia e tolleranza.

Spazi non separati. Una proprietà importante della retta R dei numeri reali consiste nel fatto che due punti diversi possono essere racchiusi in due intervalli disgiunti. Questa proprietà viene estesa alle varietà topologiche, differenziabili e analitiche: due punti distinti di una varietà si possono sempre racchiudere in due intorni aperti disgiunti. Se uno spazio topologico X gode di queste proprietà, si dice che per esso vale l'assioma di separazione di Hausdorff. Gli spazi (o le topologie) per i (le) quali questo assioma non ha luogo si dicono non-separati. Si riportano alcuni esempi.

Sia R dotato della topologia nella quale gl'insiemi aperti sono i complementi degl'insiemi finiti di punti. Questa topologia non separabile si chiama topologia di Zariski su R. Piú in generale, se X è un insieme di numeri complessi, o di punti complessi su una varietà algebrica data, ad esempio, da equazioni polinomiali in alcune variabili, gl'insiemi aperti in X secondo Zariski sono i complementi delle sottovarietà algebriche in X.

La generalizzazione della nozione di varietà algebrica è quella di schema. Uno schema è un certo spazio topologico munito di un fascio di anelli. Gli schemi piú semplici sono quelli *affini*. Se A è un qualsiasi anello commutativo con unità, lo spazio topologico dello schema affine Spec A (spettro dell'anello A) consiste in tutti gli ideali primi $P \subset A$, cioè di tutti i sottogruppi abeliani di A, chiusi rispetto al prodotto con gli elementi di A e tali che, se $xy \in P$ con $x, y \in A$, si ha $x \in P$, oppure $y \in P$. Ogni sottoinsieme $E \subseteq A$ definisce il sottoinsieme $V(E) \subseteq \mathrm{Spec}\, A$ che consiste di quei $P \subset A$ tali che $E \subset P$. I complementi $\mathrm{Spec}\, A \setminus V(E)$ formano un sistema d'insiemi aperti in un'opportuna topologia in Spec A che pure viene chiamata topologia di Zariski. Se A non ha divisori dello zero, due qualsiasi insiemi aperti in Spec A, non vuoti, hanno intersezione non vuota, cosí che Spec A non è separato. Di piú, di regola Spec A contiene punti non chiusi, piú esattamente, sottoinsiemi non chiusi costituiti da un solo punto. Cosí, se A non ha divisori dello zero, $\{0\}$ è un ideale primo in A e la chiusura del punto corrispondente in Spec A coincide con tutto lo spazio. Lo spazio Spec A può essere finito, ma dotato di una topologia non-banale come lo spazio di due punti visto precedentemente.

Infine si possono ottenere forme diverse di non-separazione «incollando» tra loro spazi topologici. Siano X, X' due spazi topologici; $U \subset X$, $U' \subset X'$ siano due loro sottoinsiemi aperti. Si supponga che esista un omeomorfismo di U su U'. Identificando nell'unione $X \cup X'$ i sottoinsiemi U e U', si ottiene un nuo-

vo spazio Y quale risultato dell'incollamento. Se, per esempio, $X=X'=\mathbf{R}$, $U=U'=\mathbf{R}\setminus\{0\}$, risulta che Y coincide ovunque con $\mathbf{R}$ ad eccezione che nello 0 e in Y il punto 0 diventa «sdoppiato»; due qualsiasi intorni di questi due punti in Y si intersecano.

La non-separazione di due punti $x, y \in X$ potrebbe essere riguardata quale modello di diversi effetti nelle scienze naturali che fino ad ora non hanno avuto una descrizione quantitativa soddisfacente. Per esempio, invece di considerare l'elettrone come una formazione puntuale, si potrebbe supporre che i punti dello spazio a distanza sufficientemente piccola diventano non-separati in una conveniente topologia. La nozione di «distanza minima» potrebbe essere allora la caratteristica quantitativa del grado di questa non-separazione rispetto all'ordinaria metrica dello spazio. La stessa cosa si potrebbe fare con un'ipotetica nozione di «durata minima».

Nelle indagini psicologiche e cosí pure nella teoria delle misurazioni l'aspetto della non-separazione interviene quale conseguenza dell'esistenza di «soglie di differenziazione», oppure di «sensibilità degli strumenti». Due punti luminosi proiettati su uno stesso fotorecettore dell'occhio diventano indistinguibili. Due oggetti materiali, che emettano luce con lungezza d'onda λ, diventano indistinguibili quando la distanza tra loro è di un ordine di grandezza minore di λ.

Zeeman, per modellare tali effetti, ha introdotto una nuova nozione matematica, gli spazi di tolleranza, i quali, a priori, non sono necessariamente dotati di una struttura topologica.

Tolleranza. Sia X un certo insieme. Si chiama tolleranza ξ su X una relazione binaria $x\xi y$ che sia riflessiva ($x\xi x$) e simmetrica (da $x\xi y$ segue che $y\xi x$). Per esempio, si può definire su $\mathbf{R}$ (o sul piano $\mathbf{R}^2$, o nello spazio $\mathbf{R}^3$) una «tolleranza con distanza minima l» ponendo $x\xi y$, se la distanza tra x e y non supera l.

Una delle proprietà interessanti della tolleranza consiste nella possibilità di ricostruire una serie di proprietà topologiche di uno spazio a partire dalla stessa tolleranza. Grosso modo, se si ha una varietà topologica con una metrica e s'introduce in essa una tolleranza con lunghezza minima l, le proprietà topologiche che «si manifestano a una distanza maggiore di l» possono essere calcolate rispetto alla tolleranza. Questo fatto è in relazione con il riconoscimento delle «forme» (*Gestalt*) in psicologia.

Zeeman ha considerato anche la possibilità di applicare questo schema alla descrizione del processo mentale. Euristicamente, si consideri lo spazio Y degli stati istantanei di tutti i neuroni del cervello; un punto di Y definisce il carattere dell'equilibrio chimico e delle frequenze delle pulsazioni di tutti i neuroni. La tolleranza viene definita con la condizione che «due stati corrispondono allo stesso pensiero». La «direzione istantanea» del pensiero è definita da un campo vettoriale v su Y; sull'insieme di questi campi è anche definito un concetto di tolleranza. Può accadere che le traiettorie di due campi, aventi inizio entrambi nei limiti della tolleranza, col tempo escano da questi limiti, come accade nei sistemi dinamici non-stabili. Si tratta dell'effetto della non-predicibilità del pensiero. Soprattutto, lo stesso aspetto del campo vettoriale $\mathbf{v}$ e la topologia del suo

quadro delle fasi variano col tempo e dipendono da eccitazioni esterne, ciò che permette di applicare l'ideologia delle catastrofi di Thom (cfr. il § 1).

Tuttavia, allo stato attuale, il modello di Zeeman non ha raggiunto un livello che permetta di dedurre predizioni quantitative atte a verifiche sperimentali.

Topologie sconnesse. In conclusione ci si soffermerà su una classe di topologie, ben studiate dai matematici, dove viene meno l'aspetto della connessione del continuo indiviso di $\mathbf{R}$. Un sottoinsieme di uno spazio topologico X si dice chiuso se il suo complemento è aperto. Uno spazio X è connesso, se esso non è l'unione di due sottoinsiemi chiusi disgiunti. Tali sono tutti gli spazi $\mathbf{R}^n$. Ogni spazio topologico è l'unione disgiunta di tutti i suoi piú grandi sottospazi connessi, cioè di tutte le sue componenti connesse. L'idea intuitiva della connessione è strettamente correlata all'idea di continuità e cosí pure alla maggioranza dei modelli fisici e ciò per la connessione degli ordinari modelli dello spazio-tempo.

Si riportano alcuni esempi caratteristici di spazi topologici completamente sconnessi, nei quali le componenti connesse sono i singoli punti. L'esempio piú semplice è lo spazio discreto, nel quale tutti i sottoinsiemi sono aperti (e chiusi). Ma la topologia può non essere cosí banale. Sia p un numero primo. Nell'anello $\mathbf{Z}$ dei numeri interi introduciamo una topologia, dichiarando due numeri interi x, y molto vicini se $x-y$ è divisibile per una grande potenza di p. In questa topologia la somma e il prodotto sono continui, ma essa non è completa; il limite di una successione di Cauchy (x_n), $x_n-x_m\to 0$ per $\min(n, m)\to\infty$, può non esistere. In modo standard è possibile completare $\mathbf{Z}$ in questa topologia e ottenere l'anello topologico completo $\mathbf{Z}_p$ dei numeri interi p-adici. Gli elementi di $\mathbf{Z}_p$ possono essere scritti sottoforma di serie:

$$\sum_{n=0}^{\infty} a_n p^n \qquad 0\leq a_n\leq p-1.$$

Le serie finite, per le quali $a_n=0$ per $n\geq n_0$, sono le rappresentazioni p-arie degli ordinari numeri interi. La topologia in $\mathbf{Z}_p$, nella quale x, y sono vicini quando $x-y$ ha la forma

$$\sum_{n=n_0}^{\infty} a_n p^n$$

con n_0 grande, si chiama topologia p-adica ed è completamente sconnessa.

Gli spazi topologici completamente sconnessi racchiudono in sé organicamente l'aspetto della continuità (topologia) e del discreto (assenza di componenti connesse non-banali). Spazi topologici piú generali possono avere componenti connesse, sui cui insiemi viene indotta una topologia non-banale, ma completamente sconnessa. Tali spazi sono molto importanti nella teoria dei numeri.

3.2. Invarianti discreti degli spazi topologici.

Sia X un certo spazio topologico. Il suo invariante discreto piú semplice è l'insieme delle sue componenti connesse, o il numero di elementi di questo in-

sieme, se esso è finito. In topologia algebrica e differenziale si è soliti usare una definizione un po' diversa di componente connessa; i punti x e y appartengono a una componente connessa per archi di X, se esiste un'applicazione continua $f : [0, 1] \to X$ tale che $f(0) = x$, $f(1) = y$, cioè se x e y possono essere uniti con un arco (cammino) in X.

Sia $\pi_0(X)$ l'insieme di tali componenti linearmente connesse di X; si tratta di quei «pezzi lineari», a due a due disgiunti, l'unione dei quali è uguale a tutto X.

Molti altri invarianti discreti di X possono essere definiti insiemi di componenti linearmente connesse di spazi topologici ausiliari che si costruiscono a partire da X. Ecco alcuni esempi.

Sia x_0 un punto prefissato (detto punto base) dello spazio X. Si chiama spazio dei cappi, e si indica con $\Omega(X, x_0)$, l'insieme di tutte le applicazioni continue $f : [0, 1] \to X$ con $f(0) = f(1) = x_0$ (archi chiusi, o cappi, che cominciano e terminano in x_0). La topologia in $\Omega(X, x_0)$ s'introduce cosí: se $K \subset [0, 1]$ è un qualsiasi sottoinsieme chiuso e $U \subset X$ è un qualsiasi sottoinsieme aperto, tutte le $f \in \Omega(X, x_0)$ con $f(K) \subseteq U$ formano un insieme aperto in $\Omega(X, x_0)$ e ogni insieme aperto è l'unione di una famiglia d'insiemi di quel tipo.

Gruppo fondamentale dello spazio X. L'insieme $\pi_1(X, x_0) = \pi_0(\Omega(X, x_0))$ si chiama gruppo fondamentale di X. In esso la legge di composizione, che lo trasforma in gruppo, è indotta dall'operazione di composizione dei cappi in $\Omega(X, x_0)$; se $f, g : [0, 1] \to X$ sono due cappi, $g \circ f$ è l'unione del cappio f, percorso durante il tempo $[0, 1/2]$ con velocità doppia, e del cappio g, percorso durante il tempo $[1/2, 1]$ con velocità doppia. L'elemento inverso alla classe f è la classe $f(1-t)$. L'elemento identico è rappresentato dal cappio $f(t) = x_0$ per ogni $t \in [0, 1]$. Il gruppo $\pi_1(X, x_0)$ in generale non è abeliano.

Pertanto il gruppo fondamentale di X definisce la quantità massimale di cappi con i quali si può «avvolgere» X, a meno di deformazioni di un cappio in un altro. Per esempio, il gruppo fondamentale di una circonferenza è $\mathbf{Z}$; al numero $m > 0$ corrisponde un cappio che corre attorno alla circonferenza m volte in una certa direzione, mentre al numero $-m$ corrisponde lo stesso cappio nella direzione contraria.

Sia dato, per esempio, nel dominio $(0, 0) \in U \subset \mathbf{R}^2$ del piano un campo vettoriale liscio $f : U \to \mathbf{R}^2$ avente un punto stazionario isolato in $(0, 0)$. Allora la funzione $\bar{f}(x) = f(x)/\|f(x)\|$ ($\|y\|$ è la lunghezza del vettore y) definisce un'applicazione di un piccolo cerchio in U con centro $(0, 0)$ sul cerchio unitario S^1 e, con ciò stesso, un elemento del gruppo $\pi_1(S^1, x_0)$ (qui la scelta di $x_0 \in S^1$ non è essenziale). Nell'identificazione $\pi_1(S^1, x_0) = \mathbf{Z}$, che richiede la scelta di un'orientazione di S^1, ad esempio il verso orario quale verso positivo, all'applicazione sopra descritta corrisponde un numero intero $i(f)$, detto indice, o numero di rotazioni del campo f nel punto corrispondente. Se il campo f corrisponde al quadro delle fasi di un sistema, $i(f)$ è un importante invariante discreto dello stato stazionario corrispondente.

Se X è una varietà differenziabile a due dimensioni compatta (da ogni suo ricoprimento se ne può estrarre uno finito), la somma degl'indici dei punti stazionari di un qualsiasi campo vettoriale su X non dipende dal campo. Questa somma è un invariante discreto di X; è la sua caratteristica di Eulero $\chi(X)$. Da qui segue che, se $\chi(X) \neq 0$, ogni campo vettoriale su X ha punti stazionari. Questo risultato si può riguardare come il fatto piú semplice della teoria qualitativa delle equazioni differenziali, che mostra l'importanza degl'invarianti topologici discreti dello spazio delle fasi. Esso è applicabile, per esempio, alla sfera a due dimensioni $S^2 : \chi(S^2) = 2$.

Superfici compatte orientate. Una varietà differenziabile X si dice orientabile se in essa vi è un atlante tale che gli jacobiani delle matrici delle funzioni di passaggio da una carta all'altra hanno tutti uno stesso segno. Un esempio di varietà non orientabile è il nastro di Möbius. Una varietà orientabile compatta X a due dimensioni è completamente definita, a meno di un omeomorfismo, dal suo gruppo fondamentale. Piú esattamente, si consideri il gruppo quoziente abeliano massimo di $\pi_1(X)$. È noto che esso è isomorfo a $\mathbf{Z}^{2g}$, dove $g \geq 0$ è un numero intero. Per ogni g esiste una superficie compatta orientabile unica, a meno di omeomorfismo, con il gruppo corrispondente; essa si dice superficie di genere g. Per $g = 0$ si ottiene la sfera; per $g = 1$ il toro; per tutti i $g \geq 1$ si può immaginare questa superficie come la sfera S^2 alla quale siano stati incollati g «manici». Incollare un manico significa ritagliare sulla sfera due piccoli fori rotondi e unire poi i loro bordi per mezzo di una superficie cilindrica.

In particolare, se il genere di X è uguale a g, si ha $\chi(X) = 2 - 2g$. Pertanto su ogni X, ad eccezione del toro, i campi vettoriali hanno punti stazionari.

Gruppi superiori di omotopia. Come sopra, X sia uno spazio topologico e si assuma $x_0 \in X$ come punto base. Per ogni $n \geq 1$ sia $\Omega^{n+1}(X, x_0) = \Omega(\Omega^n(X, x_0), \tilde{x}_0^{n+1})$, dove $\tilde{x}_0^1 = x_0$ e $\tilde{x}_0^{n+1}$ sono i cappi costanti: $f(t) = \tilde{x}_0^n$ per ogni $t \in [0, 1]$. L'insieme $\pi_0(\Omega^n(X, x_0), \tilde{x}_0^n)$, nel quale pure in modo naturale s'introduce la struttura di gruppo (per $n \geq 2$ esso è necessariamente abeliano), si chiama n-esimo gruppo di omotopia dello spazio X e si indica con $\pi_n(X, x_0)$. Questo gruppo si può pure interpretare quale gruppo delle classi dei «cappi ad n dimensioni» dello spazio X, cioè quale gruppo delle classi linearmente connesse delle applicazioni $(S^n, s_0) \xrightarrow{f} (X, x_0)$, dove $s_0 \in S^n$ è un punto base sulla sfera S^n ad n dimensioni data in R^{n+1} dall'equazione $y_1^2 + \ldots + y_{n+1}^2 = 1$.

Piú in generale, si può considerare la categoria degli spazi topologici, ognuno con un punto base, e dei morfismi, che sono classi di omotopia di applicazioni di questi spazi che applicano ogni punto base in un altro. Sia $\Pi(X, Y)$ l'insieme dei morfismi dell'oggetto X nell'oggetto Y in questa categoria. Allora $\pi_n(X, x_0)$ può essere identificato con $\Pi((S^n, s_0), (X, x_0))$ e la corrispondenza $X \to \pi_n(X, x_0)$ è un funtore rappresentabile mediante l'oggetto (S^n, s_0). Scegliendo altri oggetti in questa categoria si possono ottenere altri importanti funtori; per esempio, i gruppi di coomologia. Per ogni gruppo abeliano G e per ogni numero intero $n \geq 1$, si può costruire lo spazio $K(G, n)$, con un punto base, tale che $\Pi(X, K(G, n))$ è il gruppo $H^n(X, G)$ di coomologia a n dimensioni dello spa-

zio X con G quale gruppo dei coefficienti. Nel caso di $G = \mathbf{Z}$ e di X varietà differenziabile compatta, questi gruppi sono abeliani e hanno un numero finito di generatori. Il rango $C_n(X)$ di questi gruppi è un importante invariante numerico intero dello spazio X. Un altro invariante è la torsione di $H^n(X, G)$, cioè le caratteristiche numeriche del gruppo degli elementi di ordine finito.

I gruppi $H^n(X, G)$ e loro generalizzazioni hanno già cominciato a trovare applicazioni in fisica nello studio delle singolarità degl'integrali di Feynmann. È questo un esempio di come, a un livello di astrazione matematica molto alto, appaiano caratteristiche discrete a partire da modelli continui. [JU. I. M.].

Grünbaum, A.
1967 *Modern Science and Zeno's Paradoxes*, Wesleyan University Press, Middletown Conn.

Neumann, J. von
1958 *The Computer and the Brain*, Yale University Press, New Haven Conn. (trad. it. in V. Somenzi (a cura di), *La filosofia degli automi*, Boringhieri, Torino 1965, pp. 157-221).

Thom, R.
1971 *Stabilité structurelle et morphogénèse. Essai d'une théorie générale des modèles*, Benjamin, Reading Mass.

Dotata di una caratterizzazione gnoseologica molto generale, la categoria (cfr. **categorie/categorizzazione**) continuo/discreto si presenta in molti campi. La **fisica**, pur incompleta nelle sue conoscenze, si sforza di dare una visione universale della **natura**, fondendo l'aspetto quantistico (cfr. **quanti**) del micromondo, dominato dalla **probabilità** e dalla discontinuità, con il quadro geometrizzato dello **spazio/tempo** macroscopico, incentrato invece sulla continuità.

La descrizione dei modelli del **mondo** osservato, demandata all'elaborazione di modelli matematico-fisici (cfr. **modello**, **struttura**, **matematiche**, **assioma/postulato**), pone in evidenza vari livelli gerarchici dimensionali, partendo dalla strutturazione atomistica (cfr. **atomo e molecola**) della **materia** e dalla distribuzione energetica (cfr. **energia**) delle varie particelle (cfr. **particella**), per arrivare ai modelli della macrofisica, interpretati alla luce della teoria della **relatività** e della meccanica statistica dell'equilibrio e del non-equilibrio (cfr. **equilibrio/squilibrio**), e coinvolgenti anche i concetti di **moto** e **forza**.

Il continuo/discreto, oltre a presentarsi nei processi di **regolazione** degli organismi viventi (cfr. **organismo**), costituiti dal sistema nervoso (cfr. **cervello**) e endocrino (cfr. **integrazione**), svolge un ruolo importante nei processi di codificazione, trasmissione (cfr. **codice**, **comunicazione**) ed elaborazione dell'**informazione** ereditaria (cfr. **eredità**) degli individui, che ne assicurano lo sviluppo (cfr. **sviluppo e morfogenesi**) e la riproduzione, consentono il manifestarsi dei caratteri visibili (cfr. **gene** e **genotipo/fenotipo**) e, grazie all'intervento dei meccanismi di **mutazione/selezione** naturale, portano all'**evoluzione** del **mondo** vivente sul nostro pianeta. I problemi connessi alla trasmissione delle informazioni hanno acquistato grande rilevanza in seguito alla comparsa dei calcolatori (cfr. **analogico/digitale** e **macchina**), capaci di eseguire qualsiasi **algoritmo** (cfr. **applicazioni**).

Dipendenza/indipendenza

1. *Principî generali.*

Tutti i modelli scientifici dipendono dalla considerazione di una serie di limitazioni: alcune sfere di fatti, di fenomeni, di relazioni sono prese in esame e vengono interpretate, altre vengono considerate non essenziali, almeno in un primo momento. Ciò è condizionato, in primo luogo, dalle limitate possibilità che il cervello ha di elaborare le informazioni. La coscienza, una volta abbracciate tutte le cause e tutti gli effetti possibili, vive in un mondo del tutto illusorio.

Un altro fattore fondamentale è il carattere sociale della scienza, in quanto suo materiale grezzo possono essere soltanto fenomeni riproducibili in un senso o in un altro, e pertanto soggetti alla socializzazione. Tuttavia, un'esatta riproducibilità è impossibile, conformemente alla vecchia metafora di Eraclito sul fluire del tempo. Le leggi della meccanica quantistica che postulano il carattere essenzialmente statistico della natura, conferiscono a questa metafora una capacità ulteriore di persuasione. Una riproducibilità approssimativa presuppone che si trascurino molti fattori.

Nell'ambito di ciascun modello teorico, viene quindi postulato che le caratteristiche esaminate al suo interno siano totalmente indipendenti da tutte quelle in esso trascurate. Le stesse caratteristiche esaminate sono fra loro collegate in un sistema di relazioni, che costituiscono anche l'essenza del modello (insieme alle regole semantiche e operative d'interpretazione). Il passaggio a un altro modello fa mutare sia il sistema delle caratteristiche, sia i rapporti di dipendenza/indipendenza tra di esse. Questi ultimi hanno perciò un carattere relativo.

Si cita come semplice esempio il modello classico di Galileo: la caduta libera di un grave da un'altezza h alla superficie terrestre. Il tempo di caduta t è collegato con h dalla relazione funzionale $\sqrt{2h/g}=t$, dove g è il coefficiente «accelerazione di gravità». In questo modello il tempo di caduta non dipende dalla forma e dal peso del corpo, in quanto non si tiene conto della resistenza dell'aria. Tale dipendenza può essere presa in considerazione in un modello appena piú accurato, ma è probabile che verrà implicitamente ammessa l'indipendenza del tempo di caduta dalla posizione della Luna. Eppure quest'ultima fa variare il potenziale gravitazionale sulla superficie della Terra ed è un fattore determinante per un qualsiasi modello delle maree.

In questo articolo si adotterà quindi la seguente concezione. Le nozioni di dipendenza/indipendenza vanno annoverate fra le idee metateoretiche piú generali, con un amplissimo spettro di realizzazioni concrete. La compilazione dell'intero catalogo di questi significati concreti equivale all'esame dettagliato di tutti i modelli scientifici attualmente conosciuti: il compito, naturalmente, è irrealizzabile. Quello che ci sforziamo di fare è isolare un certo nucleo di

concetti, strettamente collegati alla descrizione della coppia dipendenza/indipendenza, all'interno di alcuni modelli fondamentali della fisica teorica. Alla fine dell'articolo sarà brevemente esaminata la loro applicazione a problemi non appartenenti al campo della fisica.

La fisica teorica utilizza la matematica quale linguaggio di lavoro: il problema che ci si è posto richiede proprio un'analisi degli strumenti matematici che esprimono i concetti di dipendenza/indipendenza, seppure nella loro correlazione con la semantica dei vari modelli. Quest'ultima proprietà del resto è fondamentale, dal momento che le stesse parole 'dipendenza/indipendenza' vengono usate in modo abbastanza limitato come termini matematici di uso corrente.

Essi vengono utilizzati in analisi e in algebra (dipendenza lineare, algebrica, funzionale), nel calcolo delle probabilità (variabili aleatorie indipendenti) e, infine, nella logica matematica (indipendenza di una data asserzione, in un linguaggio formale, da un dato sistema di assiomi).

È indubbio che l'idea di dipendenza funzionale sia uno degli strumenti principali nella descrizione matematica delle relazioni nei modelli scientifici. Il § 2 sarà interamente dedicato alla descrizione dei contesti matematici nei quali s'incontrano le parole che dànno il titolo a questo articolo. Tuttavia, l'interpretazione data all'argomento dell'articolo lo riconduce all'aspetto «realistico» della semantica della matematica. Questo significa che il concetto di dipendenza/indipendenza può esser espresso virtualmente da qualunque concetto matematico in un modello adeguato. Sembra molto istruttivo l'esame di quali sono precisamente i concetti che trovano una reale utilizzazione. Per questo motivo nell'articolo figurano, ad esempio, i prodotti tensoriali degli spazi di Hilbert, poiché essi compaiono nella descrizione dell'unione di due sistemi quantistici. Le particolari «interazioni» di particelle indistinguibili nel gas di Fermi o in quello di Bose non possono essere esattamente descritte se non dai mezzi dell'algebra multilineare. L'idea di «causalità», una delle nozioni centrali nel paradigma dei concetti fisici legati a dipendenza/indipendenza, conduce, attraverso la trasformazione di Fourier, alle funzioni analitiche e alla tecnica delle «relazioni di dispersione». L'idea di «informazione» in un sistema complesso, per esempio un organismo biologico, sembra strettamente collegata ai concetti matematici di equilibrio/squilibrio, e alle concezioni della teoria delle catastrofi di René Thom. L'idea dei vincoli imposti a un dato sistema dal mondo esterno e del carattere locale delle interazioni fra le «parti» del sistema è espressa in termini di geometria differenziale. È possibile moltiplicare il numero di questi esempi; l'analisi di alcuni di essi costituisce l'argomento del § 3. La successione dei temi di questo articolo sviluppa le corrispondenti argomentazioni dell'articolo «Applicazioni» di questa stessa *Enciclopedia*.

2. *Dipendenza/indipendenza come termini matematici.*

2.1. Relazioni nella teoria degli insiemi.

Il piú comune modello di dipendenza nell'ambito della teoria degli insiemi è descritto dalla struttura formata da n insiemi $X_1, \ldots, X_n$ e dal sottoinsieme $M \subset X_1 \times \ldots \times X_n$, detto relazione n-aria tra gli insiemi. Ad esempio, X_i può essere l'insieme dei valori a priori ammissibili per una qualche caratteristica, o parametro, del fenomeno osservato. Se tutte le n caratteristiche considerate non sono legate l'una con l'altra in alcun modo, allora il «vettore» $\langle x_1, \ldots, x_n \rangle$ dei loro valori simultanei può essere un qualunque elemento di $X_1 \times \ldots \times X_n$. Se invece tra le caratteristiche si hanno certe connessioni, allora $\langle x_1, \ldots, x_n \rangle$ può trovarsi soltanto in un qualche sottoinsieme $M \subset X_1 \times \ldots \times X_n$ che esprime tali connessioni. In questo schema s'inquadra la procedura sperimentale tipo della presentazione dei risultati di misurazioni congiunte di alcune variabili, sotto forma di punti su un grafico: ciò corrisponde al caso $X_1 = X_2 = \mathbf{R}$ (o alcuni sottoinsiemi di $\mathbf{R}$; qui è sottintesa la scelta preliminare delle unità di misura, che permettono di descrivere i valori delle caratteristiche mediante numeri reali). L'insieme delle coppie dei valori osservati $\langle x_1, x_2 \rangle$ ha di solito l'aspetto di una certa «nube di punti», la cui forma determina le caratteristiche generali della futura descrizione M del modello teorico idealizzato. Questo è il modello di partenza per l'«analisi fattoriale» nelle scienze sociali e biologiche.

Si consideri un sistema meccanico: la sua posizione istantanea può essere individuata mediante le coordinate cartesiane di un numero finito di punti distinti (ad esempio, tre punti per un corpo rigido), cioè mediante un punto opportuno nello spazio coordinato a piú dimensioni $\mathbf{R}^N$. Non è detto però che ogni punto di $\mathbf{R}^N$ corrisponda realmente ad una posizione ammissibile. L'insieme di queste ultime $M \subset \mathbf{R}^N$ è lo spazio delle configurazioni del sistema meccanico, e l'enunciazione delle sue proprietà individua i «vincoli». Ad esempio, per un corpo rigido l'insieme $M \subset \mathbf{R}^9$ è caratterizzato dalle condizioni

$$M = \left\{ (x_{ij}) \,\middle|\, \sum_{k=1}^{3} (x_{ik} - x_{jk})^2 = r_{ij}^2;\ i, j = 1, 2, 3 \right\},$$

dove x_{ik} è la k-esima coordinata dello i-esimo punto, r_{ij} è la distanza fra lo i-esimo e lo j-esimo punto.

2.2. Funzioni.

La relazione $M \subset X \times Y$ è chiamata grafico di una funzione, o di una trasformazione, da X a Y, se verifica la seguente condizione: per ogni $x \in X$ esiste un unico elemento $y \in Y$ per cui $\langle x, y \rangle \in M$. Le relazioni funzionali sono particolarmente forti: l'affermazione secondo cui il valore della caratteristica y di un dato fenomeno è completamente determinato dai valori della caratteristica x, insieme alla esplicita descrizione di questa dipendenza, rappresenta

la formulazione generale di molte leggi delle scienze naturali. Le proprietà generali delle funzioni e una serie di modelli che utilizzano dipendenze funzionali sono state descritte nell'articolo «Applicazioni». Si ritornerà piú avanti sugli aspetti specifici delle dipendenze funzionali importanti per il tema svolto in questa sede.

2.3. La dipendenza in algebra.

Si riportano alcuni esempi caratteristici dell'uso del termine 'dipendenza' in ambiti algebrici. Lo schema generale è il seguente: in $X_1, \ldots, X_n$ sono assegnate alcune leggi di composizione; con il loro aiuto è costruita la famiglia di trasformazioni $f_i : X_1 \times \ldots \times X_n \to Y_i$, agli insiemi Y_i appartengono determinati elementi $y_i \in Y_i$; la dipendenza di un dato tipo è determinata dalla relazione $R = \bigcup_i f_i^{-1}(y_i)$. Ecco i principali casi particolari.

Dipendenza lineare. Gli elementi $x_1, \ldots, x_n$ di uno spazio vettoriale X sul campo k sono detti linearmente dipendenti se esistono elementi del campo $a_1, \ldots, a_n \in k$, non tutti nulli, tali che $\sum_{i=1}^{n} a_i x_i = 0$. Altrimenti, $x_1, \ldots, x_n$ sono linearmente indipendenti: ciò significa che dall'uguaglianza $\sum_{i=1}^{n} a_i x_i = 0$ segue che tutti gli a_i sono nulli. La riformulazione nei termini dello schema sopra descritto è la seguente: la famiglia di trasformazioni $f : X \times \ldots \times X \to X$ si compone di trasformazioni lineari non nulle del tipo $f(x_1, \ldots, x_n) = \sum a_i x_i$, $a_i \in k$; le relazioni di dipendenza lineare $M \subset X \times \ldots \times X$ sono $\bigcup_f f^{-1}(0)$. In seguito, ovviamente, non si ridurranno a questo schema le definizioni formulate in modo abituale.

Il numero massimo dei vettori linearmente indipendenti nello spazio è detto dimensione dello spazio vettoriale X (la dimensione è infinita se tale numero non è finito). Questo è l'esempio piú semplice d'invariante discreto di una struttura matematica, descritto in termini di dipendenza. Le sue diverse varianti, anche in casi non-lineari, sono dette dimensione, rango, ecc.

La dimensione dello spazio fisico è uguale a tre, quella dello spazio-tempo è uguale a quattro. Ciò significa che per la perfetta localizzazione di un evento elementare, accaduto in un punto dello spazio-tempo, sono sufficienti quattro numeri reali: tale condizione comporta fortissime limitazioni sulla forma delle leggi della natura.

Si supponga che in uno spazio lineare di dimensione finita X sia scelta una base, cioè X s'identifichi con lo spazio k^m delle m-ple di elementi del campo k. Allora si applica il seguente metodo per stabilire la dipendenza o l'indipendenza lineare di n vettori assegnati $x_1, \ldots, x_n \in X$. Si scrivano le coordinate di questi vettori come righe di una matrice (con n righe e m colonne). Se $n > m$, i vettori $x_1, \ldots, x_n$ sono linearmente dipendenti. Se $n \leq m$ i vettori sono linearmente indipendenti se, e solo se, tale matrice A contiene una sottomatrice di dimensione $n \times n$ formata da n colonne della matrice A, il cui determinante («minore» di A) è diverso da zero. Si ricordi che il determinante $\det(a_{ij})$ di una matrice quadrata su k è l'espressione $\sum_s \varepsilon(s) a_{1s(1)} \ldots a_{ns(n)}$, dove s descrive tutte le permutazioni dell'insieme $(1, \ldots, n)$, e $\varepsilon(s) = \pm 1$ è il segno della permutazione s. Si osservi che per $k = \mathbf{R}$, qualora lo spazio $\mathbf{R}^n$ sia munito della solita struttura euclidea, il determinante della matrice $n \times n$ è uguale al volume orientato del parallelepipedo $\left\{ \sum_{i=1}^{n} \alpha_i x_i \,\middle|\, 0 \leq \alpha_i \leq 1 \right\}$, sotteso dai vettori riga di questa matrice.

Spesso la definizione di dipendenza lineare è espressa nella seguente forma non-simmetrica: il vettore x_n è detto linearmente dipendente dai vettori $x_1, \ldots, x_{n-1}$ se esiste una relazione del tipo $x_n = \sum_{i=1}^{n-1} a_i x_i$, $a_i \in k$, oppure, il che è lo stesso, una relazione del tipo $\sum_{i=1}^{n} a_i x_i = 0$ con $a_n \neq 0$.

Dipendenza algebrica e funzionale. Sia K un certo campo, e $k \subset K$ un certo suo sottocampo. Gli elementi $x_1, \ldots, x_n \in K$ sono detti algebricamente dipendenti su k se esiste un polinomio non identicamente nullo $F \in k\,[T_1, \ldots, T_n]$ con coefficienti in k tale che $F(x_1, \ldots, x_n) = 0$ nel campo K.

Il numero massimo di elementi algebricamente indipendenti del campo K su k è detto grado di trascendenza di K su k. Sia ad esempio X una varietà compatta complessa, sulla quale le funzioni meromorfe formano il campo K. X può essere ad esempio una varietà proiettiva algebrica irriducibile. Allora il grado di trascendenza di K sui numeri complessi $\mathbf{C}$ non è superiore a, e nel caso algebrico coincide con, la dimensione n della varietà X concepita nel seguente modo: quasi tutti i punti di X possiedono un intorno analiticamente isomorfo a $\mathbf{C}^n$.

Per ogni campo k esiste un'estensione massimale $\bar{k} \supset k$, avente su k un grado di trascendenza nullo. Essa è univocamente definita a meno di un isomorfismo su k, e è detta completamento algebrico di k. Tutti i suoi elementi si ottengono come radici di polinomi non identicamente nulli in una variabile con coefficienti in k. Il completamento algebrico del campo dei numeri razionali $\mathbf{Q}$ può essere realizzato come il sottocampo dei numeri complessi $\mathbf{C}$, formato da tutti i numeri algebrici $\bar{\mathbf{Q}}$. Tuttavia, il grado di trascendenza di $\mathbf{C}$ su $\mathbf{Q}$ è infinito.

Il concetto di dipendenza algebrica comporta una serie di utili varianti. Per la teoria dei numeri e per la geometria algebrica è importante il seguente concetto di dipendenza intera. Siano $A \subset B$ due anelli commutativi con unità. L'elemento $x \in B$ è detto intero su A se soddisfa l'equazione $F(x) = 0$, dove $F \in A\,[T]$ è un polinomio con coefficienti in A e il cui coefficiente della potenza di grado massimo è l'unità.

Per la teoria delle funzioni e per la geometria differenziale e analitica è importante il concetto di dipendenza delle funzioni che ha il seguente significato. Siano $f_1(x), \ldots, f_n(x)$ funzioni definite sulla varietà X. Se esse verificano una relazione del tipo $F(f_1, \ldots, f_n) = 0$, allora sono dette dipendenti. Imponendo

a f e soprattutto a F la condizione di appartenenza all'una o all'altra classe di funzioni, se ne ricavano i concetti di dipendenza algebrica, analitica, lineare, e cosí via. Se $f_1, \ldots, f_n$ sono funzioni differenziabili di n variabili reali $x_1, \ldots, x_n$, vale il seguente importante criterio di assenza di qualsiasi relazione non banale $F(f_1, \ldots, f_n) = 0$ con la funzione differenziabile F. Tale criterio afferma che il determinante della matrice $(\partial f_i/\partial x_j)$, il cosiddetto jacobiano, risulta diverso da zero in un certo punto. Ciò significa che, se le funzioni f_i, con un'approssimazione lineare vicino ad un certo punto, sono linearmente indipendenti, allora esse sono anche funzionalmente indipendenti.

Qualsiasi funzione regolare, definita su una varietà differenziabile n-dimensionale, dipende in modo regolare dalle n funzioni coordinate. In questo modo, la dimensione geometrica assume un aspetto funzionale, essendo espressa mediante il numero massimo possibile di funzioni indipendenti. Questo stesso schema è usato nella geometria algebrica ed analitica.

Moduli e gruppi abeliani. Sia M un modulo su un anello A. Gli elementi $x_1, \ldots, x_n \in M$ sono detti indipendenti se dalla relazione

$$\sum_{i=1}^{n} a_i x_i = 0$$

con $a_i \in A$ segue che tutti gli a_i sono uguali a zero. Questa definizione ripete testualmente quella di dipendenza lineare negli spazi vettoriali. Tuttavia appaiono nuovi importanti fenomeni se l'anello A non è un campo, dal momento che non tutti i moduli (neppure in condizioni di finitezza) sono isomorfi al modulo delle n-ple A^n, $n \geq 0$. Ad esempio, nel modulo $\mathbf{Z}/m\mathbf{Z}$ sull'anello dei numeri interi ($m \neq 0$ numero intero) qualsiasi famiglia di elementi è dipendente, comprese tutte le famiglie ad un solo elemento $\{x\}$. I moduli, isomorfi ad A^n, sono detti liberi di rango n, e uno strumento potente d'indagine di un qualunque modulo consiste nel confrontarlo con i moduli liberi.

Se si suppone che M sia finitamente generato, ciò significa che esiste un omomorfismo suriettivo $A^{n_1} \xrightarrow{f_1} M$. Con M_1 è designato il suo nucleo. Esso può non essere finitamente generato, ma lo è nell'importante classe degli anelli noetheriani. Si può allora definire un omomorfismo suriettivo $A^{n_2} \xrightarrow{f_2} M_1$, scegliendo i generatori di M_1. Si può quindi applicare questo procedimento al nucleo M_2 dell'omomorfismo f_2, e cosí via. Si ottiene in tal modo la successione di moduli e di omomorfismi

$$\ldots \longrightarrow A^{n_i} \xrightarrow{f_i} A^{n_{i-1}} \xrightarrow{f_{i-1}} \ldots \longrightarrow A^{n_1} \xrightarrow{f_1} M \longrightarrow 0$$

che forma una successione esatta.

Gli elementi del nucleo di f_1 possono essere interpretati come relazioni lineari, o sizigie, fra i generatori del modulo M scelti per la costruzione dell'omomorfismo f_1; gli elementi del nucleo di f_1 possono essere interpretati come relazioni fra queste relazioni, e via dicendo. Fu Hilbert a studiare per la prima volta una tale successione di sizigie o, per dirla in termini moderni, una tale risoluzione del modulo M; egli dimostrò il teorema fondamentale in

base al quale essa, per l'anello dei polinomi $k[T_1, \ldots, T_n]$ in n variabili su un campo, termina in n passi, nel senso che il nucleo f_n sarà certamente un modulo libero. Con questo teorema Hilbert ha posto la base della moderna algebra omologica, il cui problema tipico consiste nello studio di quegli invarianti del modulo M che è possibile ricavare dall'esame di una sua qualunque risoluzione libera. In particolare, la lunghezza di tale risoluzione è una caratteristica tipica della dimensione (dell'anello e del modulo). Il termine 'sizigie' per le relazioni lineari risale al linguaggio degli astronomi: i rapporti numerici tra i periodi di rivoluzione dei pianeti conducono al ripetersi regolare delle configurazioni delle loro mutue posizioni, ad esempio le opposizioni.

I moduli sull'anello $\mathbf{Z}$ sono semplicemente dei gruppi abeliani. Ogni gruppo abeliano G, nella notazione additiva, ammette una risoluzione (complesso esatto) del tipo $0 \to \mathbf{Z}^n \to \mathbf{Z}^m \to G \to 0$. Il numero $m-n$ non dipende dalla scelta della risoluzione e si chiama il rango r del gruppo G. Un gruppo di rango r è isomorfo alla somma diretta $\mathbf{Z}^r \oplus G_0$, dove G_0 è finito. In un gruppo di rango r esiste un sistema di r elementi indipendenti, ma $r+1$ elementi arbitrari sono già dipendenti.

È conveniente descrivere il concetto fondamentale di «dimensione (scalare) di una grandezza fisica», in termini di dipendenza e di rango di un gruppo abeliano.

In un qualunque modello fisico figurano concetti tipo tempo, lunghezza, massa; forza, lavoro e potenza; carica elettrica, ecc.

La loro introduzione è accompagnata dalla descrizione di procedure sperimentali per la scelta di unità di misura e per l'espressione numerica delle corrispondenti caratteristiche del sistema in termini del dato sistema di unità. In ogni modello alcune unità vengono considerate fondamentali e indipendenti, altre invece derivate da queste. È molto naturale considerare le unità derivate come elementi del gruppo abeliano libero generato dalle unità fondamentali. Cosí nella fisica newtoniana le unità di lunghezza L, di tempo T e di massa M sono considerate le unità fondamentali. Allora le unità di velocità LT^{-1}, di forza MLT^{-2} e di lavoro ML^2T^{-2} sono considerate derivate (come è usuale, le si scrive con notazione moltiplicativa; cfr. le notazioni piú classiche del tipo cm/sec per la velocità). Un elemento del gruppo abeliano libero, generato dalle unità fondamentali, corrisponde a una data grandezza fisica e si chiama dimensione di questa grandezza. È naturale ritenere che il gruppo dimensionale stesso D non dipenda dalla scelta delle unità: l'elemento L corrisponde alla lunghezza indipendentemente dalla convenzione usata per la scelta dell'unità di lunghezza.

Il gruppo dimensionale D è una caratteristica molto importante del modello fisico. Lo sviluppo delle scienze naturali è stato accompagnato da due processi opposti che hanno cambiato d'improvviso la nostra concezione della natura: da una parte l'ampliamento del gruppo D relativo ai modelli fondamentali, in connessione con la scoperta di variabili di nuova natura, dall'altra la riduzione di questo gruppo in connessione con la scoperta di relazioni tra le dimensioni che prima erano considerate indipendenti.

Cosí, la prima indagine sulla natura dell'elettricità è stata accompagnata dall'aggiunta, alle unità L, T, M, dell'unità q di carica elettrica (nel sistema moderno CGSE la carica elementare dell'elettrone è uguale a $4{,}80 \cdot 10^{-10}$ u. e. s.). D'altra parte, l'accettare l'idea della costanza della velocità della luce nei sistemi inerziali conduce, nella teoria della relatività speciale, al legame $L = T$ nel gruppo corrispondente D: ogni scelta dell'unità di lunghezza permette di assumere, come unità di tempo, il tempo necessario alla luce per percorrere tale unità di lunghezza. (Per la misura delle distanze cosmiche ci si serve dell'unità «anno-luce»; in tale definizione, ovviamente, è utilizzata la procedura inversa della determinazione della lunghezza attraverso il tempo).

Ogni scoperta di un'unità naturale di misura per questa o quella grandezza fisica conduce perciò alla scoperta di una nuova relazione nel gruppo D usato in precedenza, e alla sua sostituzione con il gruppo quoziente corrispondente a questa relazione. La costante gravitazionale di Newton G, la velocità della luce c, la costante di Planck $\hbar$ (che fornisce l'unità universale per l'azione) e la carica dell'elettrone e rappresentano le grandi svolte di questo cammino. Nel 1899 Planck poté indicare il sistema universale di unità fondamentali

$$L^* = (\hbar G/c^3)^{1/2} = 1{,}6 \cdot 10^{-33} \text{ cm}$$
$$T^* = (\hbar G/c^5)^{1/2} = 5{,}3 \cdot 10^{-44} \text{ sec}$$
$$M^* = (\hbar c/G)^{1/2} = 2{,}2 \cdot 10^{-5} \text{ g}$$

che riducono il gruppo D a uno banale e permettono di considerare tutte le grandezze della fisica come costanti adimensionali. È degno di nota che le espressioni costruite indipendentemente dalle unità fondamentali (compresa la carica dell'elettrone) hanno la stessa dimensione anche in un gruppo piú grande, cosicché i loro rapporti sono sempre costanti numeriche universali. La piú famosa di esse è la costante di struttura fine $e^2/\hbar c$: la velocità media dell'elettrone nell'atomo dell'idrogeno è grosso modo pari a $1/137$ (della velocità della luce). Attualmente questo numero viene determinato con grande esattezza in modo sperimentale. La scoperta di un modello teorico, in cui il suo significato potrebbe risultare derivato da principî fondamentali, è considerata uno dei piú interessanti compiti della microfisica.

La considerazione del gruppo D per un arbitrario modello fisico pone notevoli limitazioni alla forma matematica delle relazioni in questo modello: ogni relazione rappresentata, ad esempio, da un rapporto di polinomi fra i parametri, deve essere omogenea, cioè tutti i monomi devono avere la stessa dimensione (tenendo conto della dimensione delle costanti). In espressioni tipo $\exp \varphi$, φ deve essere «adimensionale». Ad esempio, nella formulazione della meccanica quantistica di Feynmann figura l'espressione $\exp(iS/\hbar)$ dove S è un'azione classica; evidentemente, essendo $S/\hbar$ una grandezza adimensionale, l'azione è misurabile in unità universali di Planck.

Le considerazioni di dimensionalità a volte limitano talmente la forma delle relazioni possibili da permettere di determinarle in modo quasi univoco. Esse sono anche largamente usate nella costruzione di modelli sperimentali, ad

esempio nell'idrodinamica per determinare i coefficienti di similitudine con l'aiuto dei quali sono ricalcolate le misurazioni dei modelli per ottenere le informazioni sui fenomeni in tempo reale.

L'esistenza di unità universali di misura è strettamente legata all'inesplicabile fatto dell'indistinguibilità delle particelle elementari di un tipo, ad esempio, di elettroni. Come mostreremo piú avanti, lo stesso fatto di questa indistinguibilità, insieme ai fondamenti quanto-meccanici della descrizione dei sistemi di particelle, conduce ad effetti molto peculiari delle interazioni, che non possono essere spiegati sulla base delle idee classiche relative alle «forze», ad esempio, di attrazione o di repulsione.

Il seguente esempio curioso è ripreso da Weisskopf [1970]: «L'altezza delle montagne è determinata da costanti fisiche fondamentali». Si tenga presente quanto segue: la cima piú alta della Terra, l'Everest, ha un'altezza di circa 10 km; perché non ci sono montagne piú alte? Anche senza tener conto dei meccanismi di distruzione e di erosione geologica, si dimostra che la natura fisica della materia solida delle rocce della Terra, le dimensioni della Terra e i valori delle costanti fondamentali limitano la possibile altezza nell'ordine di alcune decine di chilometri. Una montagna di massa M e di altezza h può avere il peso MG con il risultato di «fondersi» o di «liquefarsi» nella sua parte inferiore fino ad un'altezza x. L'energia di fusione, in questo modo, determina un'altezza massima fino alla quale non ha ancora luogo la fusione. Tenendo conto dell'ordine delle grandezze corrispondenti, si ottiene la disuguaglianza

$$\frac{h}{a_0} \leq \beta\eta \frac{\alpha}{\alpha_G} \frac{1}{N^{1/3}} \frac{1}{A^{5/3}}$$

dove a_0 è il raggio dell'atomo dell'idrogeno, $\alpha_G = Gm_p^2/\hbar c \sim 5{,}9 \cdot 10^{-39}$ (m_p è la massa del protone), $\alpha = e^2/\hbar c \simeq 1/137$ è la costante di struttura fine, $A \approx 60$ è il peso atomico medio della materia della montagna, $N \sim 3 \cdot 10^{51}$ è il numero dei protoni e dei neutroni nella materia della Terra, $\beta \sim 0{,}1$, $\eta \sim 0{,}2$ sono le caratteristiche del calore di fusione.

La dipendenza nei gruppi non-commutativi. Siano G un gruppo non necessariamente commutativo e $x_1, \ldots, x_n$ suoi elementi. Essi sono detti dipendenti, se esiste un'espressione non banale uguale all'unità del tipo $x_{a_1}^{\varepsilon_1} x_{a_2}^{\varepsilon_2} \ldots x_{a_m}^{\varepsilon_m} = 1$. Qui $a_i \in \{1, \ldots, n\}$, $\varepsilon_i = \pm 1$, e la non-banalità significa che non esiste un indice i per cui $a_i = a_{i+1}$ e $\varepsilon_i = -\varepsilon_{i+1}$, cioè non si possono ridurre le moltiplicazioni presenti. Una definizione piú invariante è la seguente. Per ogni numero intero $n \geq 1$ è possibile determinare un gruppo libero G_n di rango n con generatori liberi $\{y_1, \ldots, y_n\}$. Questo gruppo è definito completamente dal fatto che un qualunque omomorfismo $G_n \xrightarrow{f} G$ in un qualunque gruppo G è definito univocamente da $f(y_1), \ldots, f(y_n)$ ed è possibile scegliere queste immagini arbitrariamente. Se $f(y_i) = x_i$, il nucleo dell'omomorfismo f è detto insieme delle relazioni tra gli elementi $x_1, \ldots, x_n$ del gruppo G. Questi elementi sono indipendenti se, e solo se, il nucleo f si compone solo dell'unità, cioè se f è iniettivo. Una definizione piú generale di struttura algebrica libera di

un dato tipo, generata dall'insieme $\{y_1, ..., y_n\}$ può essere data in termini di funtori, in connessione col funtore «dimenticante della struttura algebrica».

Il rango di un gruppo libero ha un'interessante proprietà che lo distingue, ad esempio, dalla dimensione lineare. Quest'ultima può soltanto diminuire con il passaggio a un sottospazio e a uno spazio-quoziente. In un gruppo libero di rango due, esistono sottogruppi liberi di rango finito arbitrario e persino numerabile. Questa è una delle piú tipiche manifestazioni degli effetti speciali della non-commutatività nell'algebra.

Il concetto di gruppo libero permette d'introdurre un'importante costruzione: la rappresentazione del gruppo mediante generatori e relazioni.

Sia I un insieme arbitrario. Allora si può costruire un gruppo $G(I)$ contenente I e detto gruppo libero con insieme di generatori I, che possiede la seguente proprietà: per qualunque gruppo G avente una famiglia di elementi $\{x_i \mid i \in I, x_i \in G\}$ esiste un unico omomorfismo di gruppi $f : G(I) \to G$ tale che $f(i) = x_i$. Il gruppo $G(I)$ è detto libero con l'insieme libero di generatori I. Esso è definito in modo unico. Il gruppo G_n sopra considerato è $G(y_1, ..., y_n)$ nelle nuove notazioni.

Ogni elemento diverso dall'unità del gruppo $G(I)$ si presenta semplicemente nella forma $i_1^{a_1} \ldots i_k^{a_k}$, dove $i_j \in I$, a_j sono numeri interi, $i_j \neq i_{j+1}$ per ogni j.

Qualunque sottoinsieme di elementi $\mathcal{J}$ del gruppo G permette di considerare in G due sottogruppi: $G_{\mathcal{J}}$ e $N_{\mathcal{J}}$. Qui $G_{\mathcal{J}}$ è il piú piccolo sottogruppo di G, che contiene $\mathcal{J}$, cioè è l'intersezione di tutti i sottogruppi contenenti $\mathcal{J}$, mentre $N_{\mathcal{J}}$ è il piú piccolo divisore normale di G che contiene $\mathcal{J}$. L'insieme $\mathcal{J}$ è detto sistema di generatori del gruppo G se $G_{\mathcal{J}} = G$. Se $\mathcal{J}$ è un sistema di generatori del gruppo G, allora ogni applicazione dell'insieme I su $\mathcal{J}$ determina un'applicazione suriettiva $G(I) \to G$ e definisce cosí un isomorfismo del gruppo G con il gruppo-quoziente di $G(I)$ rispetto a un suo divisore normale opportuno.

Sia ora $K \subset G(I)$ un insieme di elementi del gruppo libero, $N_K \subset G(I)$ il divisore normale da esso generato. È d'uso dire che il gruppo quoziente $G(I)/N_K$ è dato dall'insieme di generatori I e di relazioni K tra essi.

Molti gruppi concreti, ad esempio i gruppi fondamentali degli spazi topologici, sono naturalmente dati mediante generatori e relazioni. Il seguente problema è detto «il problema dell'identità delle parole» nei gruppi: ammesso che siano dati I e K (ad esempio, siano insiemi finiti dati mediante un elenco), come si fa a sapere se due elementi $x, y \in G(I)$ determinano il medesimo elemento nel gruppo quoziente $G(I)/N_K$? Il problema dell'identità delle parole è detto decidibile con un algoritmo (per i dati I, K) se esiste un algoritmo che, applicato a qualunque coppia $x, y \in G(I)$ dà risposta alla domanda se xy^{-1} si trovi in N_K.

Novikov diede per primo esempi di gruppi con I, K finiti, per i quali il problema dell'identità delle parole non è decidibile algoritmicamente. Basandosi sul lavoro successivo di Highmann, non è difficile costruire un esempio semplice di tale gruppo con I finito, ma con K infinito; argomentazioni piú sottili permettono di modificare la costruzione di questo esempio in modo da ottenere anche K finito.

Per la spiegazione di questo esempio basta ricordare che è possibile determinare un sottoinsieme E di numeri naturali, numerabile ma non decidibile. La numerabilità di E significa che esiste un algoritmo per l'enumerazione di tutti gli elementi di E in un certo ordine. La non-decidibilità di E significa che non esiste un algoritmo che, una volta applicato a un qualunque numero naturale, risponda alla domanda se questo numero si trovi in E. Si costruisca ora il gruppo con i quattro generatori $\{a, b, c, d\}$ e con le relazioni $b^{-m}ab^m = d^{-m}cd^m$ per tutti gli $m \in E$. Non è difficile verificare che in questo gruppo la relazione $b^{-x}ab^x = d^{-x}cd^x$ $(x \geq 1)$ è realizzata nel caso, e solo nel caso, in cui $x \in E$. Dalla definizione di E segue allora che anche per tali coppie di parole il problema dell'identità risulta algoritmicamente non-decidibile.

In questo modo si ha la seguente fondamentale condizione: può accadere di conoscere dall'inizio tutte le relazioni N_K tra i generatori di un certo gruppo (come conseguenza delle relazioni date K), ma non si è in grado di determinare in modo effettivo se un dato elemento di $G(I)$ appartiene all'insieme delle relazioni N_K. Certamente, questo è un caso particolare dei fenomeni generali di indecidibilità delle teorie matematiche, scoperti per la prima volta da Gödel.

In due parole, questi fenomeni consistono nel fatto che in ogni teoria matematica formale sufficientemente ricca (nella quale cioè è inclusa la descrizione dell'aritmetica) esistono sempre delle proposizioni che non dipendono in essa da nessun sistema assegnato di assiomi, cioè non possono essere né dimostrate né confutate.

Si esamineranno ora piú dettagliatamente questi concetti d'indipendenza logica e le loro implicazioni filosofiche.

2.4. Logica.

La dipendenza logica è definita nel contesto seguente: viene esaminato un certo linguaggio formale con prefissate regole di formazione degli enunciati e con regole di deduzione, cioè di produzione di enunciati a partire da altri. Un certo sistema di enunciati $\mathcal{E}$ è inizialmente ammesso; si tratta dei cosiddetti assiomi. Tutti gli enunciati che si possono ricavare da $\mathcal{E}$ mediante le regole di deduzione costituiscono la teoria formale T. L'enunciato P è detto indipendente da $\mathcal{E}$ se né P né la negazione di P non appartengono a $\mathcal{E}$. È noto che esistono linguaggi e sistemi di assiomi talmente ricchi da poter essere considerati come modelli formali adeguati di quasi tutta la matematica, in relazione alle possibilità di deduzione in matematica.

D'altra parte, la matematica ha anche un aspetto ricco di contenuto che si esprime con la rappresentazione non-formale della verità di un enunciato matematico in questa o in un'altra interpretazione nella teoria («elementare») degl'insiemi. Come la formulazione degli assiomi e le dimostrazioni dei teoremi in matematica sono in ultima analisi procedure finite per registrare ed elaborare dei testi di lunghezza finita, cosí l'idea astratta di verità si appella di regola all'idea di una serie infinita, numerabile o anche non numerabile, di verifiche. Si ammette che la congettura di Fermat («se $x^n + y^n + z^n = 0$;

x, y, z, n interi; $n \geq 3$; allora $xyz = 0$») è vera o falsa dal momento che si può immaginare un'enumerazione successiva dell'insieme infinito delle quaterne (x, y, z, n). Per ognuna delle quaterne la verità o la falsità dell'ipotesi di Fermat è verificata in modo finito. Se consideriamo potenzialmente conclusa questa successione di verifiche, allora essa deve contenere in sé la prova o la confutazione di tale ipotesi.

Per le applicazioni delle scienze naturali sono essenziali entrambi gli aspetti della conoscenza matematica. La deduzione di un teorema non evidente da postulati plausibili, e la sua successiva applicazione alla chiarificazione di fenomeni fisici, costituisce il lavoro quotidiano del fisico teorico. Con non minore ragione, tuttavia, si può dire che al fisico interessa la verità del teorema piú che la sua dimostrazione, benché la sua idea di verità si distingua molto dalla descrizione precedente per essere basata sulla realtà fisica.

Il progresso della logica matematica nel XX secolo ha messo in luce il profondo abisso esistente tra i concetti di «verità» e di «deducibilità da assiomi dati»: l'ultima affermazione è di gran lunga piú povera, qualunque sistema di assiomi si venga ad adottare. I matematici si sono già abituati all'esistenza di «problemi non risolvibili» all'interno della loro scienza, cioè di affermazioni, indipendenti dai piú forti sistemi di assiomi, che sono ammessi o hanno la possibilità di essere ammessi dalla maggioranza degli scienziati. Le conseguenze di questa scoperta per la «scienza applicata» o, piuttosto, per la filosofia della scienza, sono comprese in modo assai meno soddisfacente. Non si sa come porsi davanti alla possibilità che in un qualche modello matematico, per esempio nella chimica quantistica, la spiegazione del comportamento osservato di una macromolecola possa consistere nell'ipotetica verità di un qualche teorema matematico, che risulterà formalmente non-dimostrabile.

Questo problema, forse, non è troppo importante dal punto di vista pratico. Il calcolo effettivo di sistemi complessi esige spesso una tale mole di lavoro per cui le limitazioni sono dovute alla capacità delle macchine e alla brevità della vita piuttosto che a difficoltà logiche. Si è disposti ad accontentarsi di un modello semplificato o trasformato, di un breve riassunto di calcoli approssimativi, dell'intuizione fisica e di criteri pratici di diverso genere che assicurino l'efficacia del modello.

Tuttavia, il filosofo della scienza non può trascurare le circostanze indicate esaminando i problemi di riducibilità. La chimica e la biologia si riducono alla fisica? Ponendo delle domande di tal genere, bisogna anzitutto rendersi conto che si tratta sempre di un modello teorico delle leggi della fisica, della chimica e della biologia, e non direttamente della natura. In secondo luogo, all'interno di qualunque modello teorico che in ultima analisi è destinato ad essere matematizzato, le conoscenze sottintendono forse una dipendenza logica o un'astrazione del tipo della verità matematica?

L'attento esame delle diverse varianti del significato che può essere attribuito al problema della riducibilità prova che esso probabilmente è stato mal posto e si serve di caratteri di astrattezza per i quali accettare questa o quella risposta è piú un atto di fede che di conoscenza scientifica. I ragionamenti nei

quali viene proposto d'immaginare un operatore di Schrödinger per un sistema biologico e poi di tirare le conclusioni sul suo possibile comportamento, prendendo lo spunto da semplici postulati generali della meccanica quantistica, sono troppo grossolani. Un qualunque sostenitore della «forza vitale» potrebbe ad esempio dichiarare che nello spazio di Hilbert, che corrisponde a un sistema vivo, esiste uno speciale operatore hermitiano dell'«entelechia», che non c'è negli spazi meno complessi dei sistemi fisici, e «riconciliare» in questo caso i sostenitori della riducibilità con i loro avversari. D'altra parte, la possibilità perennemente incombente dell'indipendenza logica delle leggi della vita dalle leggi della natura non vivente nei nostri modelli matematici, lascia ampio spazio alle speculazioni di qualunque genere.

Va sottolineato ancora una volta, richiamandosi ai modelli matematici, che le astrazioni della verità e le astrazioni della deducibilità, problemi di riduzionismo in ultima analisi, sembrano strettamente legati ai problemi della finitezza dei mezzi di descrizione forniti da una qualunque lingua, in opposizione alle astrazioni dell'infinità che entrano forzatamente nella semantica della lingua se essa ha pretese di efficienza. Senza prendere posizione nei confronti di questo dualismo, il filosofo non può nemmeno porre chiaramente la questione della riducibilità di principio di alcuni modelli della natura ad altri, senza già parlare dei tentativi ammissibili di soluzione di tale questione.

2.5. Il calcolo delle probabilità. Lo spazio di probabilità.

Il calcolo delle probabilità come disciplina applicata si occupa degli «eventi aleatori». Non saranno qui esaminate le questioni filosofiche che sono legate al problema della casualità. Nei modelli matematici di casualità di «tipo classico» l'effetto di casualità deriva dalla mancanza d'informazione sulle condizioni in cui accade un evento, o dalla sua eccessiva sovrabbondanza che costringe a servirsi solo di una sua piccola parte. Sono soggetti ad esame non gli eventi individuali, ma soltanto le loro caratteristiche medie rispetto a un insieme reale o ideale di eventi aleatori di ugual tipo, cioè rispetto a un «insieme statistico». Nei modelli della fisica quantistica la casualità è interpretata come insita nella natura delle cose: praticamente tutti i valori delle osservabili microfisiche sono aleatori, e hanno carattere deterministico soltanto le affermazioni sulle probabilità di questi valori. In questo paragrafo ci si occuperà soltanto dei modelli di tipo classico.

Il punto di partenza della teoria matematica è uno spazio con misura (Ω, μ), cioè uno spazio di eventi elementari o spazio di probabilità. L'assegnazione di un punto di Ω è interpretata come informazione completa su un evento elementare. Assegnare un sottoinsieme misurabile $Y \subset \Omega$ – un «evento» – significa dare un'informazione non completa che distingue gli eventi elementari di una qualche classe. La misura $\mu(Y)$ degli eventi è interpretata come la sua probabilità. Per questo motivo la misura μ è supposta non-negativa, e $\mu(\Omega) = 1$.

2.6. Esempio: le prove di Bernoulli.

Si consideri lo spazio $\Omega_n = \{0, 1\} \times \ldots \times \{0, 1\}$ (n volte). Ogni suo punto $\omega \in \Omega$ è una successione con elementi zero o uno di lunghezza n; tutto lo spazio si compone di 2^n elementi. Scegliamo i numeri $p, q \geq 0$, $p+q=1$ e definiamo la misura $\mu = \mu_{p,q}$ sulla σ-algebra di tutti i sottoinsiemi di Ω, ponendo $\mu(\omega) = p^{|\omega|} q^{n-|\omega|}$, dove $|\omega|$ è il numero degli uno nella successione ω (e quindi $n-|\omega|$ è il numero di zeri in essa). Non è difficile verificare che $\mu(\Omega_n) = \sum_{\omega \in \Omega_n} \mu(\omega) = 1$, cosicché (Ω_n, μ) è uno spazio di probabilità. Si tratta di un modello per il seguente esperimento: il lancio n volte di seguito di una moneta non simmetrica con la probabilità, uguale a p, di avere testa, e uguale a q di avere croce.

2.7. Esempio: lo spazio di Maxwell.

Sia N un numero intero. Si consideri un certo volume di un gas ideale monoatomico formato da N molecole di massa m. L'essere ideale significa che le molecole non interagiscono; l'essere monoatomico garantisce l'assenza di gradi di libertà interni della molecola, su alcuni dei quali può concentrarsi una parte dell'intera energia del gas E.

La configurazione istantanea del gas verrà descritta con un punto dello spazio euclideo di dimensione $3N$ delle velocità di tutte le molecole: $\omega = (v_{ix}, v_{iy}, v_{iz} \mid i = 1, \ldots, N)$, v_{ix} è la proiezione della velocità sull'asse delle x, ecc. Lo spazio delle posizioni è la sfera Ω di dimensione $(3N-1)$:

$$\sum_{i=1}^{N} (v_{ix}^2 + v_{iy}^2 + v_{iz}^2) = \frac{2E}{m}.$$

Se la temperatura del gas è uguale a T, allora l'energia media di una molecola è $3kT/2$ (k è una costante universale), e il raggio di Ω è $(3kN/m)^{1/2}$. Da considerazioni di simmetria deriva che la misura di probabilità su Ω coincide con l'usuale misura di Lebesgue, normalizzata con la condizione $\mu(\Omega) = 1$. Si considerino gli eventi consistenti nel fatto che la proiezione v_{ix} è compresa tra i valori α e β. La misura della corrispondente calotta sferica per $N \to \infty$ tende a

$$\sqrt{\frac{m}{2\pi kT}} \int_{\alpha}^{\beta} \exp\left(-\frac{mx^2}{2kT}\right) dx.$$

Questa è la legge di distribuzione della velocità secondo Maxwell.

2.8. Variabili aleatorie e distribuzioni.

Una funzione misurabile $\xi : \Omega \to \mathbf{R}^n$ è detta variabile aleatoria vettoriale. La misurabilità significa che per ogni sottoinsieme di Borel $A \subset \mathbf{R}^n$ la controimmagine $\xi^{-1}(A)$ è misurabile rispetto alla misura μ su Ω; gli insiemi di Borel

sono gli elementi della più piccola σ-algebra, in $\mathcal{P}(\mathbf{R}^n)$, comprendente tutti i parallelepipedi.

Legge (o distribuzione) della variabile aleatoria ξ è la misura μ_ξ su $\mathbf{R}^n$, per la quale $\mu_\xi(B) = \mu(\xi^{-1}(B))$, cioè $\mu_\xi(B)$ è la probabilità che i valori di ξ si trovino in B.

2.9. Indipendenza degli eventi e delle variabili aleatorie.

Gli eventi $A, B \subset \Omega$, sono detti indipendenti se $\mu(A \cap B) = \mu(A)\mu(B)$. Le variabili aleatorie $\xi_1 : \Omega \to \mathbf{R}^{n_1}$, $\xi_2 : \Omega \to \mathbf{R}^{n_2}$ sono dette indipendenti se per ogni $A_i \subset \mathbf{R}^{n_i}$ gli eventi $\xi_1^{-1}(A_1)$, $\xi_2^{-1}(A_2)$ sono indipendenti. In modo analogo si può definire il concetto di famiglia di variabili aleatorie totalmente indipendenti.

L'insieme delle variabili aleatorie su uno spazio di probabilità (Ω, μ) è un oggetto di regola «non osservabile» e tanto grande quanto lo spazio stesso. Il legame del modello con l'esperimento si realizza di solito utilizzando un piccolo numero di funzionali sulle variabili aleatorie. Fra i più semplici e i più importanti di essi sono annoverate la speranza matematica e la varianza.

Si dice speranza matematica della variabile aleatoria scalare $\xi : \Omega \to \mathbf{R}$ il «valore medio» $M\xi = \int_{-\infty}^{+\infty} \xi \, d\mu_\xi(x)$. Lo scarto quadratico medio dalla speranza matematica viene detto varianza $D\xi$:

$$D\xi = M(\xi - M\xi)^2 = \int_{-\infty}^{+\infty} (\xi - M\xi)^2 \, d\mu_\xi(x).$$

Il numero $\operatorname{cov}(\xi, \eta) = M(\xi - M\xi)(\eta - M\eta)$ è detto covarianza delle due variabili aleatorie $\xi, \eta : \Omega \to \mathbf{R}$. Dalle definizioni segue facilmente che, se ξ, η sono indipendenti, allora $\operatorname{cov}(\xi, \eta) = 0$. Per questo nel caso generale $\operatorname{cov}(\xi, \eta)$ è la misura numerica più semplice di dipendenza (sarebbe meglio dire di «non-indipendenza») delle variabili aleatorie ξ, η. Giacché, come non è difficile vedere, $D(\xi + \eta) = D\xi + D\eta + 2 \operatorname{cov}(\xi, \eta)$, la varianza della somma di variabili aleatorie indipendenti è uguale alla somma delle loro varianze.

Sia ad esempio (Ω_n, μ) lo spazio di Bernoulli. Si scelgano due sottoinsiemi $I, \mathcal{J} \in \{1, \ldots, n\}$, $I \cap \mathcal{J} = \emptyset$, due funzioni $f : I \to \{0, 1\}$ e $g : \mathcal{J} \to \{0, 1\}$, e si definiscano A, B come gli eventi, formati dai punti ω, per i quali le restrizioni di ω su $I, \mathcal{J}$ coincidono rispettivamente con f, g. Questi eventi, come si può verificare, sono indipendenti. Intuitivamente ciò è in accordo col fatto che sapere che sono usciti testa o croce durante gli esperimenti di I non dice niente sui risultati degli esperimenti di $\mathcal{J}$.

Moltissimi modelli probabilistici discendono dall'ipotesi che le caratteristiche per noi interessanti del fenomeno sono condizionate dall'azione di un gran numero di fattori aleatori indipendenti. Da ciò deriva che molti teoremi sono formulati come asserzioni sul comportamento asintotico di funzionali relativi a un numero crescente di variabili aleatorie indipendenti. La deduzione della distribuzione di Maxwell, in sostanza, si era già servita di quest'idea. Si darà ancora qualche esempio.

2.10. Legge dei grandi numeri.

La si cita nella formulazione piú semplice di Čebyšev. Si supponga che su (Ω, μ) sia data una successione di variabili aleatorie a due a due indipendenti $\xi_1, \ldots, \xi_n, \ldots$ con uguale speranza matematica a e con varianze uniformemente limitate: $D\xi_i < c$ per tutti gli i. Si ponga

$$\eta_n = \frac{1}{n}\sum_{i=1}^{n} \xi_i$$

e si indichi con $p_n(\varepsilon)$ la misura dell'insieme per cui $|\xi_n - a| > \varepsilon$. Il teorema di Čebyšev dice che, qualunque sia $\varepsilon > 0$, $p_n(\varepsilon) \to 0$ per $n \to \infty$. In altre parole, se N è sufficientemente grande, allora il valore medio di (ξ_n) per $n \leq N$ con probabilità quasi 1, differisce arbitrariamente poco da a: le deviazioni accidentali da a in caso di addizione hanno la tendenza a crescere molto piú lentamente del numero dei termini addizionati.

2.11. Il teorema di Poisson.

Si consideri la successione degli spazi di Bernoulli $(\Omega_n, \mu_{p_n, q_n})$, $n = 1, 2, 3, \ldots$ Sia $|\omega|_n$ la variabile aleatoria $\omega \to |\omega|$ su Ω_n. Supposto che per $n \to \infty$ risulti $p_n \to 0$ in modo che esista il limite $\lim_{n\to\infty} np_n = \lambda$, allora per ogni $k = 0, 1, 2, \ldots$ la probabilità dell'evento $|\omega|_n = k$ tende, per $n \to \infty$, a $(\lambda^k/k!)e^{-\lambda}$.

Lo spazio dei numeri interi non negativi $\Omega = \{0, 1, 2, \ldots\}$ con la misura $\mu(k) = (\lambda^k/k!)e^{-\lambda}$ si chiama spazio di Poisson. Piú in generale, la variabile aleatoria $\xi : \Omega \to \{0, 1, 2, 3, \ldots\}$, per la quale $\mu_\xi(k) = (\lambda^k/k!)e^{-\lambda}$, si chiama variabile aleatoria con distribuzione di Poisson.

Intuitivamente essa descrive la distribuzione del numero delle prove «riuscite» nelle serie lunghe, nelle quali la probabilità di successo (o, piuttosto, d'insuccesso) è limitata in modo inversamente proporzionale alla lunghezza della serie e le prove sono indipendenti. In realtà si può ammettere una dipendenza non troppo forte delle prove senza cambiare le conclusioni del teorema di Poisson. Tuttavia una forte dipendenza, cioè la «contagiosità del caso», viola questa deduzione. Si hanno tali situazioni quando la messa fuori servizio di elementi isolati di un sistema grande provoca immediatamente la messa fuori servizio degli elementi legati ad essi, o quando è studiata la statistica delle malattie infettive.

2.12. Teorema di Moivre-Laplace.

Esso ha a che fare con la successione degli spazi di Bernoulli $(\Omega_n, \mu_{p,q})$, dove p e q non dipendono da n. Si considerino, come prima, le variabili aleatorie $|\omega|_n : \omega \to |\omega|$. In virtú della legge dei grandi numeri, $(|\omega|_n/n) \to p$ se $n \to \infty$ con probabilità uno. Il teorema di Moivre-Laplace dimostra che la deviazione

$|\omega|_n$ dal valore np cresce proporzionalmente a $\sqrt{n}$. Piú precisamente, s'introduca la variabile aleatoria $\nu_n(\omega) = (|\omega|_n - np)/\sqrt{npq}$ su Ω_n. Allora per ogni $x \in \mathbf{R}$ e per $n \to \infty$ la probabilità che $\nu_n > x$ tende alla funzione

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}dy.$$

La variabile aleatoria ξ con distribuzione

$$\mu_\xi(A) = \frac{1}{\sqrt{2\pi}}\int_A e^{-x^2/2}dx$$

è detta normale (con i parametri (0, 1)). Si tratta di una distribuzione notevole, che svolge un ruolo importante nell'intero spettro dei modelli probabilistici.

Una delle cause di ciò è spiegata dal cosiddetto teorema limite centrale di Laplace, il quale mostra che la conclusione del teorema di Moivre-Laplace resta vera se c'interessa la deviazione dalla media della somma di variabili aleatorie totalmente indipendenti con deboli limitazioni sulle loro distribuzioni (si osservi che $|\omega|_n$ è la somma $\sum_{i=1}^{n} \xi_i(\omega)$, dove ξ_i è la i-esima coordinata del punto ω).

2.13. Teorema limite centrale.

È conosciuto in molte varianti: si formulerà qui la variante di Ljapunov.

Sia $\xi_1, \xi_2, \ldots$ una successione di variabili aleatorie indipendenti su uno spazio di probabilità e si supponga che per ogni k esistano le grandezze $M\xi_k = a_k$, $D\xi_k = \sigma_k^2$ e $M|\xi_k - a_k|^3$. Posto poi

$$B_n^2 = \sum_{k=1}^{n} \sigma_k^2,$$

si supponga che per $n \to \infty$ si abbia

$$B_n^{-3}\sum_{k=1}^{n} M|\xi_k - a_k|^3 \to 0$$

(condizione di Ljapunov). Si indichi con ν_n la variabile aleatoria

$$\nu_n = B_n^{-1}\left(\sum_{i=1}^{n} (\xi_i - a_i)\right).$$

In questo caso la probabilità dell'avvenimento $\nu_n < x$, $x \in \mathbf{R}$, tende verso $(1/\sqrt{2\pi})\int_{-\infty}^{x} e^{-x^2/2}dx$ uniformemente rispetto a x. In altre parole, la distribuzione limite della variabile aleatoria ν_n è normale con parametri (0, 1).

Questo teorema è uno degli strumenti fondamentali della teoria degli errori delle osservazioni, dove si suppone che la grandezza registrata in un esperimento sia la somma del suo «vero» valore e di un errore casuale di osservazione. Allora ripetere la sua registrazione (in una serie di osservazioni in uguali condizioni) e quindi mediare i risultati, conduce con grande probabilità al

valore piú vicino a quello vero, perché la media della somma degli errori possibili in n osservazioni diminuirà per esempio come $1/\sqrt{n}$ (se $\sigma_k = \sigma$ non dipende da k).

La stessa legge della radice quadrata opera nella teoria della «passeggiata aleatoria». L'esempio classico è il moto browniano: in un tempo t la deviazione media della particella dalla posizione iniziale crescerà come $\sqrt{t}$. Ciò è condizionato dall'indipendenza della successione degli urti aleatori subiti dalla particella come conseguenza dell'agitazione termica delle molecole.

Nella teoria delle comunicazioni, ad esempio delle trasmissioni di segnali su distanze cosmiche, riveste un ruolo importante il problema della separazione di un segnale debole sullo sfondo di un forte rumore. Nella variante piú semplice esso può essere risolto se ogni segnale viene ripetuto molte volte periodicamente, e se nell'apparecchio radio viene sommato il segnale ricevuto con il rumore su molti periodi. Allora l'ampiezza della componente utile crescerà proporzionalmente al numero dei periodi, ma l'ampiezza del rumore crescerà proporzionalmente soltanto alla radice quadrata di questo numero, cosicché si potrà separare il segnale dal rumore dopo un numero sufficiente di ripetizioni. Qui, senza dubbio, si suppone anche che il carattere del rumore in diversi periodi sia aleatorio e indipendente, cosicché essi daranno luogo ad interferenze.

3. *Sistemi chiusi e interazioni.*

3.1. Generalità.

Fra tutte le astrazioni della scienza moderna, che esprimono l'idea di dipendenza/indipendenza, quella centrale è, probabilmente, la concezione di sistema chiuso. Si suppone che esistano, o che possano essere create artificialmente o, infine, che si possano immaginare in un esperimento mentale, delle parti dell'universo, la cui evoluzione, in un determinato periodo della loro esistenza, avviene secondo loro proprie leggi interne. In questo periodo il mondo esterno o non ha in generale interazione con il sistema (il sistema è isolato oppure è libero), oppure quest'interazione può essere descritta sommariamente e il suo effetto è incluso nelle equazioni che definiscono la struttura e l'evoluzione del sistema («vincoli», «parametri esterni», ecc.).

Un'astrazione di tal genere è sottintesa quando la fisica parla di «volume di un gas a una data temperatura e pressione», di «atomo dell'idrogeno», di «sistema solare» o di «campo elettromagnetico nel vuoto».

L'analisi di un sistema complesso ha lo scopo d'individuare quei sottosistemi potenzialmente chiusi e piú semplici. La *sintesi* ricostruisce la rete delle loro effettive relazioni, delle loro interazioni, ecc.

L'idea di sistema chiuso presenta diversi aspetti. L'aspetto della limitatezza della sua esistenza nel tempo conduce a mettere in evidenza le «condizioni iniziali», cioè le condizioni al momento della sua nascita come sistema chiuso, oppure le sue caratteristiche nel momento in cui cominciamo ad osservarlo. (Questo momento può riferirsi, nei modelli matematici, a un passato infinitamente remoto, cosa che non cambia il nocciolo della questione, infatti «l'infinitamente tanto tempo fa» viene ad essere semplicemente «molto tempo fa in confronto con la durata temporale caratteristica della vita del sistema». Negli esperimenti sulla dispersione delle particelle elementari, «molto tempo fa» può significare «frazioni di microsecondi fa», e il tempo caratteristico della vita può essere dell'ordine di 10^{-10} secondi e meno). L'aspetto di limitatezza di un sistema chiuso nello spazio può essere espresso inserendo la descrizione sommaria della sua interazione col resto dell'universo sotto forma di assegnazione delle «condizioni al contorno»: sulla superficie di un grave è costante una data temperatura; sull'oscillatore agisce una forza conosciuta; il campo si annulla all'infinito.

In parte nella meccanica classica, ma ancora di piú in quella quantistica, è essenziale l'aspetto duale della localizzazione del sistema, non nella rappresentazione spazio-temporale, ma in quella impulsivo-energetica. Ad esempio, si può parlare in modo convenzionale della localizzazione spazio-temporale di un fotone, soltanto nei momenti del suo assorbimento o della sua diffusione. Al di fuori di questi momenti il fotone possiede soltanto un'energia finita e un impulso. Questo è solo uno degli aspetti della complessità del significato di sistema chiuso nella meccanica quantistica.

La separazione delle «leggi della natura», cioè delle leggi dello sviluppo dei sistemi chiusi, dalle condizioni iniziali e al contorno, è alla base dei moderni principî della descrizione scientifica del mondo. Le leggi della natura sono scritte nella lingua delle formule matematiche e costituiscono il supporto discreto, finito, esplicito della conoscenza scientifica. Le condizioni iniziali e al contorno si riferiscono alla componente infinita, continua, variabile. Ora, chiaramente, l'idea matematica del continuo non può essere espressa in ultima analisi nella sua necessaria pienezza; le sue interpretazioni operazionali non sono meno deboli, poiché l'idea del continuo si basa sulle astrazioni della divisibilità illimitata e (potenzialmente) della misurazione arbitrariamente esatta.

A questo punto è naturale accostarsi al terzo aspetto, forse il meno evidente, della presentazione dei sistemi chiusi. Esso consiste nell'osservare che il concetto di sistema chiuso implica inevitabilmente la precisazione del limite del dettaglio con il quale il sistema stesso viene esaminato.

Il sistema solare come oggetto della meccanica celeste viene di regola analizzato non solo come isolato dalla parte restante della nostra galassia, ma inoltre i pianeti vengono ritenuti punti materiali, provvisti di una massa, ma né di volume né definiti da una struttura fisico-chimica. Il campo elettromagnetico nell'approssimazione classica si basa sul modello di una coppia di vettori degli sforzi, variabili con continuità da un punto all'altro dello spazio e da un momento all'altro del tempo, senza che si tenga conto dei fenomeni quanto-meccanici. Il volume di un gas come oggetto macroscopico può essere descritto dalla sua equazione di stato indipendentemente dalla rappresentazione

atomica. Per dirla in breve, se anche con grossolana approssimazione, si considera un sistema fisico isolato dal restante mondo infinito (oppure molto grande), esso appare ancora un mondo infinitamente complesso, qualora si decida di descriverlo in tutti i dettagli. Ogni sua immagine matematica trascura quei particolari che sono caratterizzati da sufficientemente piccole caratteristiche spazio-temporali e/o impulso-energetiche. Le piccole fluttuazioni si compensano; non verranno presi in considerazione piccoli spostamenti energetici; ecc. ecc. Perciò è possibile dire cosa verrà trascurato e cosa non verrà preso in considerazione, soltanto in rapporto a un modello ulteriore piú esatto del sistema; allora può risultare che gli effetti che dànno solo piccole correzioni alle caratteristiche di cui ci si è precedentemente interessati, svolgano tuttavia un ruolo determinante nel quadro del mondo come lo conosciamo noi. La scoperta dello spin dell'elettrone spiegava dapprima soltanto piccolissime deviazioni di un fascio di elettroni negli esperimenti del tipo di quelli di Stern-Gerlach, o la decomposizione appena percettibile delle linee spettrali. Tuttavia, dopo la scoperta del legame dello spin con la statistica (principio di Pauli), risultò che l'esistenza di atomi stabili e, in ultima analisi, di corpi materiali, è assicurata dalle proprietà di questa sorprendente «osservabile non-classica». Qui si accennerà soltanto brevemente al problema dell'esistenza di diversi livelli di descrizione di questo o quel sistema, quando due «modelli vicini» differiscono nella loro struttura in modo cosí forte come quello classico e quello quantistico. Si osserverà soltanto che la relatività dei concetti di «dipendenza/indipendenza» si manifesta con particolare evidenza per l'appunto in congiunzione con tali modelli.

Forse questo aspetto di «rozzezza della descrizione» di un sistema chiuso è piú evidente nei modelli cosmologici. Oggetto del modello cosmologico è «tutto l'universo in ogni tempo». Anche non analizzando esempi concreti, non è difficile immaginarsi che un cosí immenso «sistema chiuso» può diventare oggetto dell'osservazione fisica soltanto dopo l'astrazione da quasi tutte le caratteristiche dell'universo con le quali ha a che fare quasi tutta l'altra scienza.

3.2. I sistemi classici e il principio della minima azione.

Nella descrizione matematica dei piú svariati sistemi isolati ha un ruolo fondamentale uno schema generale per la deduzione delle equazioni dell'evoluzione: il principio della minima azione.

Ecco un breve ed evidente esempio: s'immagini un punto materiale – cioè una «pallina» che può muoversi in una scanalatura curva su di un piano verticale, in virtú della forza di gravità – e si trascurino sia l'attrito sia la rotazione della pallina. La descrizione matematica di questo sistema è costituita dai seguenti oggetti:

a) Spazio delle configurazioni (spazio degli stati) del sistema. È l'insieme di tutte le possibili posizioni istantanee della pallina. Queste ultime possono essere rappresentate con un numero reale q, cioè con una coordinata che, tuttavia, può essere introdotta in molti modi differenti: ad esempio, calcolando la

distanza della pallina da un punto prefissato lungo l'orizzontale o lungo la scanalatura, oppure modificando la sua coordinata angolare relativamente a un punto opportuno. Il sistema possiede un grado di libertà; il suo movimento è unidimensionale; la scanalatura rappresenta il «vincolo» imposto alla pallina.

b) Spazio delle fasi del sistema. È l'insieme delle coppie $(q, \dot{q})$, dove q è la coordinata della posizione istantanea della pallina, e $\dot{q}$ è la rapidità della sua variazione nel tempo. Durante le diverse traiettorie del movimento della pallina essa può passare da un punto a un altro con diversa velocità, per questo le coppie di valori $(q, \dot{q})$ possono descrivere tutto il piano (o una sua parte, se la lunghezza della scanalatura è finita e/o se le velocità per un qualunque motivo sono limitate). Teoricamente si potrebbe considerare anche lo spazio delle possibili terne $(q, \dot{q}, \ddot{q})$ (coordinata, velocità, accelerazione), ma nei sistemi meccanici di tale tipo tutta la traiettoria reale del sistema è definita semplicemente dalla posizione e dalla velocità iniziali: l'accelerazione è localmente definita da queste ultime, e quindi l'integrazione delle equazioni differenziali del movimento fornisce tutta la traiettoria. Nei sistemi piú complessi lo spazio delle fasi può includere particolari gradi di libertà per varie derivate d'ordine superiore.

Quadro delle fasi del sistema. Si tratta delle curve nello spazio delle fasi, rappresentanti $(q(t), \dot{q}(t))$ per tutti i movimenti reali possibili della pallina, per $t \in \mathbf{R}$. Queste curve sono dette anche traiettorie nello spazio delle fasi. Nella descrizione nello spazio delle fasi possono esservi traiettorie formate da un solo punto: la posizione di equilibrio della pallina; in esse la velocità è uguale a zero, e la coordinata q corrisponde alle posizioni di «buche» e di «sommità» della scanalatura. La pallina nella buca si trova in equilibrio stabile: se la si allontana un po' verso la parete e la si lascia andare, incomincia a rotolare da una parte all'altra, rimanendo tutto il tempo vicino al fondo. La pallina sulla sommità si trova in equilibrio instabile: se la si allontana un po' verso la parete e la si lascia andare, rotola in giú. Sono inoltre possibili dei casi degeneri: i tratti orizzontali della scanalatura e i punti di flesso dove la scanalatura a sinistra sale, e a destra scende, ma in quel punto la tangente è orizzontale. Se la forma della scanalatura è descritta da una funzione regolare, allora tutti i punti di equilibrio sono gli *estremi* dell'altezza della scanalatura come funzioni di una coordinata orizzontale: per piccoli spostamenti lungo l'orizzontale la parte lineare della variazione di altezza è uguale a zero.

In fisica è piú ragionevole lavorare non con una scanalatura alta, ma con l'energia potenziale della pallina $V(q)$ nel campo di gravità (calcolata a partire da un prefissato livello orizzontale). Nelle posizioni di equilibrio $V(q)$ è estremale.

Piccole oscillazioni. Le traiettorie generali del movimento della pallina possono avere forme abbastanza complicate, ma esiste un'importante classe di movimenti che ha una struttura molto semplice: piccole oscillazioni intorno alle posizioni di equilibrio stabile. Se q è scelta opportunamente e $\dot{q} = q = 0$ nella posizione di equilibrio, allora il movimento è vicino all'oscillazione di

un oscillatore armonico: $\ddot{q}+k^2q=0$, $q=A\cos k(t-t_0)$. È necessario soltanto che $V(q)$ in prossimità di $q=0$ abbia la forma di una parabola $V(q)\sim q^2$ a meno di termini d'ordine superiore; un adeguato piccolo cambiamento di q la trasformerà allora in una parabola esatta, ma una opportuna modificazione di $\dot{q}$ renderà il movimento armonico. Le piccole oscillazioni rappresentano uno dei processi fisici piú stabili; con ciò si spiega in parte il ruolo degli oscillatori armonici negli altri processi fisici.

Leggi di conservazione. Sia E l'energia cinetica e V l'energia potenziale della pallina: si tratta di funzioni sullo spazio delle fasi. Allora su ogni traiettoria la somma $E+V$ è conservata: questa è la legge di conservazione dell'energia.

Introduzione delle interazioni con il mondo esterno. A rigor di termini, sin dal principio il sistema non era isolato: sulla pallina agiva il campo esterno della gravità. Tuttavia, lo si è potuto descrivere come isolato, ritenendo questa azione conosciuta in anticipo e non dipendente in alcun modo dal comportamento della pallina stessa: è sostanziale per l'appunto quest'ultimo fatto. Inoltre, il campo di gravità è potenziale e non può compiere lavoro in un ciclo chiuso del moto del sistema.

Inoltre si sarebbe potuto prendere in considerazione l'attrito della pallina con la scanalatura: esso produce dissipazione di energia (per trasformazione in energia termica) e in fin dei conti piccole oscillazioni smorzate; questo è ancora uno dei motivi per cui le piccole oscillazioni sono importanti.

Si sarebbero potute includere anche una fonte esterna di energia, ad esempio un campo magnetico variabile, che metta in moto la pallina (di ferro) o una variazione della forma della scanalatura a causa di un influsso esterno. Nell'articolo «Continuo/discreto» di questa stessa *Enciclopedia* si erano osservati i risultati «catastrofici» sul movimento della pallina che in tal modo possono verificarsi: ad esempio, la caduta in una piccola buca creatasi in un lato.

Commento. Nonostante tutta la semplicità di quest'esempio, il complesso dei principali concetti sopra descritti si adatta a molti sistemi chiusi piú complicati. Ogni modello matematico presuppone una piú o meno evidente introduzione dello spazio degli stati. Le coordinate generalizzate in questo spazio possono essere indipendenti oppure soddisfare relazioni del tipo «vincoli». La prima e fondamentale complessità del sistema consiste nella dimensione di questo spazio: il numero dei gradi di libertà del sistema. Nei sistemi meccanici esso è finito, ma rifiutare il principio delle «azioni a distanza» richiede l'introduzione dei campi; il «valore istantaneo del campo» è una funzione in una regione dello spazio, cioè l'insieme infinito dei suoi valori in tutti i punti della regione: in questo senso lo spazio dei valori del campo ha dimensione infinita. Il modello di un gas come insieme di particelle classiche in una scatola ha, a rigor di termini, dimensione finita, ma la sua dimensione è cosí grande (numero di Avogadro $\approx 10^{23}$) che è «praticamente infinita».

Le posizioni di equilibrio del sistema possono avere un ruolo molto importante e spesso sono definite come punti estremali di un opportuno funzionale sullo spazio degli stati o sullo spazio delle fasi.

Le leggi di conservazione delle diverse grandezze possono esprimere in modo molto evidente l'«isolamento» del sistema. Se vi è un numero sufficiente di grandezze che si conservano, già lo stesso fatto della loro esistenza permette di ricevere un'informazione molto importante sulle proprietà dell'evoluzione del sistema.

Traiettorie virtuali e azione. Si dà ora una descrizione piú precisa degli schemi matematici corrispondenti. In essa ha un ruolo importante un principio, profondo e non evidente, delle scienze naturali, secondo il quale il movimento di un sistema può essere spesso descritto come la sua «posizione di equilibrio» in un senso generalizzato del termine. Con piú precisione, si considerano i punti dello spazio degli stati della pallina nella scanalatura come «stati virtuali di equilibrio» del sistema; tra di essi gli stati effettivi di equilibrio sono individuati come punti estremali della funzione «energia potenziale». In modo analogo è possibile considerare tutte le possibili curve parametrizzate $(q(t), \dot{q}(t))$, $t_0 \leq t \leq t_1$ nello spazio delle fasi come «traiettorie virtuali» della pallina e tentare d'individuare tra di esse le traiettorie effettive come estremali per un certo funzionale. Tale funzionale esiste effettivamente: è detto azione e viene scritto $\int_{t_0}^{t_1} L(q(t), \dot{q}(t), t)\,dt$, dove L è un'opportuna funzione (nel caso della pallina nella scanalatura è la differenza fra la sua energia cinetica e quella potenziale). Si tratta del cosiddetto principio della minima azione (o, piú precisamente, dell'azione estremale).

Piú in generale, a ogni sistema chiuso di una determinata classe si fa corrispondere una struttura matematica (E, S) formata dall'insieme E delle «traiettorie virtuali» del sistema e dalla funzione $S : E \to \mathbf{R}$ a valori reali, che possiede fisicamente la dimensione di un lavoro e il cui valore è detto azione della traiettoria virtuale data. La classe delle traiettorie reali del sistema viene scelta con la condizione che l'azione su di esse ammetta un valore massimale, per lo meno in un senso locale. L'insieme E si realizza di solito come l'insieme delle sezioni locali di un fibrato differenziabile. Le coordinate locali in tale fibrato sono dette coordinate generalizzate del sistema. L'azione è data di regola dall'integrale di una forma differenziale che dipende dalle coordinate generalizzate del sistema e dalle loro derivate sulla base del fibrato: questa forma è detta densità lagrangiana del sistema. La condizione di minimalità dell'azione è sostituita dall'annullamento della sua variazione prima nel senso del calcolo delle variazioni, che conduce alle classiche equazioni di Eulero-Lagrange. Si dànno qui di seguito alcuni esempi.

Sistemi meccanici. Sia N una varietà differenziabile, i punti della quale sono interpretati come posizioni istantanee del sistema. (Generalizzazione dell'esempio iniziale). L'insieme delle traiettorie virtuali E è l'insieme delle immagini delle applicazioni $\mathbf{R} \to N$ definite sui segmenti $[t_0, t_1]$ della retta del tempo $\mathbf{R}$ o, il che è lo stesso, l'insieme delle sezioni locali della proiezione

$N \times \mathbf{R} \to \mathbf{R}$. Queste immagini sono supposte differenziabili il numero necessario di volte.

Siano $(q_1, \ldots, q_n)$ delle coordinate locali su N che ricoprono completamente una traiettoria virtuale $(q_1(t), \ldots, q_n(t))$, $t_0 \leq t \leq t_1$. Allora l'azione in essa è definita da un integrale del tipo

$$\int_{t_0}^{t_1} L\left(q_i, \frac{dq_i}{dt}, \ldots, \frac{d^n q_i}{dt^n}, t\right) dt.$$

La funzione $L(q_i, \dot{q}_i, \ddot{q}_i)$, dove tutte le q_i e le derivate sono considerate variabili indipendenti, è detta lagrangiana del sistema, o densità lagrangiana. In molti modelli $n = 1$, e per semplicità ci si limiterà a questo caso.

Come è stato detto, le traiettorie reali si distinguono, tra quelle virtuali, per la condizione di estremalità dell'azione. Per la deduzione delle equazioni differenziali del moto questa condizione viene un po' modificata e, seguendo Eulero e Lagrange, viene sostituita con l'estremalità locale dell'azione nel seguente senso. Si consideri la traiettoria $q_i(t)$ e una sua piccola variazione $(q_i(t) + \delta q_i(t))$, dove le funzioni δq_i sono diverse da zero esattamente all'interno di $[t_0, t_1]$. Nell'approssimazione lineare rispetto a δq_i la variazione dell'azione ha la forma

$$S(q+\delta q) - S(q) \simeq \int_{t_0}^{t_1} \sum_i \left[\frac{\partial}{\partial q_i} L(q)\, \delta q_i + \frac{\partial}{\partial \dot{q}_i} L(q)\, \delta \dot{q}_i\right] dt =$$

$$= \int_{t_0}^{t_1} \sum_i \left(\frac{\partial L}{\partial q_i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i}\right) \delta q_i\, dt.$$

L'ultima espressione è ottenuta integrando per parti il termine $(\partial L/\partial \dot{q}_i)\, \delta \dot{q}_i\, dt$, e tenendo conto del fatto che $\delta q_i(t_0) = \delta q_i(t_1) = 0$. Sotto ipotesi analitiche naturali, servendosi dell'arbitrarietà degl'incrementi δq_i, è possibile da qui dedurre che lungo le traiettorie reali

$$\frac{\partial L}{\partial q_i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} = 0.$$

Queste sono le equazioni differenziali di Eulero-Lagrange. Eccettuato un insieme singolare, esse determinano il movimento del sistema quando siano assegnate le sue coordinate generalizzate (q_i) e le velocità generalizzate $(\dot{q}_i)$ nell'istante iniziale t_0.

Osservazioni complementari. Dal punto di vista matematico è meglio lavorare senza le coordinate locali nel sistema delle definizioni fondamentali. È possibile fare ciò se si associa ad ogni fibrato regolare $N \to M$ la successione di fibrati regolari $\ldots \to \mathcal{J}^k N \to \mathcal{J}^{k-1} N \to \ldots \to \ldots \to \mathcal{J}^0 N = N \to M$ chiamati fibrati dei getti. Un punto di $\mathcal{J}^k N$ corrisponde a una classe di sezioni locali del fibrato $N \to M$, che hanno un contatto con ordine di tangenza non inferiore a k in un dato punto di N. Le densità lagrangiane sono definite in tale ambito come forme differenziali esterne su un $\mathcal{J}^k N$, aventi dimensione massima «lungo la base» di M e dimensione nulla «lungo le fibre». L'integrale della restrizione di una tale forma a una sezione locale è l'azione di questa sezione rispetto alla data densità lagrangiana.

Il formalismo variazionale diventa, in questa variante, un nuovo capitolo assai interessante della geometria differenziale.

Un sistema di N punti materiali interagenti nello spazio-tempo galileiano è di solito descritto con il seguente modello. Le coordinate generalizzate $\{q_{i_1}, q_{i_2}, q_{i_3}\}$ sono le $3N$ coordinate delle posizioni dei punti $i = 1, \ldots, N$, in un sistema di coordinate cartesiane. Posto

$$v_i^2 = \sum_{j=1}^{3} \dot{q}_{ij}^2,$$

allora la lagrangiana del sistema ha la forma

$$L = \sum_{i=1}^{N} \frac{m_i v_i^2}{2} - V(q_{ij}).$$

Qui m_i è la massa del punto i-esimo, $V(q_{ij})$ è una funzione delle posizioni dei punti. Il termine

$$T = \sum_{i=1}^{N} \frac{m_i v_i^2}{2}$$

viene detto energia cinetica del sistema, V è la sua energia potenziale.

La descrizione dell'evoluzione mediante la lagrangiana e mediante il sistema delle equazioni di Eulero-Lagrange si applica non soltanto a sistemi isolati nell'accezione piú ristretta del termine, ma anche a sistemi che hanno interazioni con un altro sistema se l'evoluzione di quest'altro sistema può ritenersi assegnata, e se è possibile trascurare l'influenza reciproca esercitata su di essa dal primo sistema. Caso tipico è il movimento di un sistema di punti materiali in un campo esterno, gravitazionale o elettromagnetico, che è prodotto da oggetti (come nell'esempio della pallina) con massa molto grande (a confronto del sistema).

In matematica l'influsso del sistema esterno si manifesta sulla forma della lagrangiana nel modo seguente. Immaginiamo che il nostro sistema chiuso I e il sistema II che agisce su di esso siano riuniti in un sistema generale isolato I+II, la lagrangiana del quale ha l'aspetto $L(q_{\mathrm{I}}, \dot{q}_{\mathrm{I}}; q_{\mathrm{II}}, \dot{q}_{\mathrm{II}})$. Ritenendo q_{II}, $\dot{q}_{\mathrm{II}}$ funzioni conosciute del tempo, possiamo sostituirle in L. Variando q_{I} e $\dot{q}_{\mathrm{I}}$, le equazioni di Eulero-Lagrange per il sistema I avranno formalmente l'aspetto di prima, ma L dipenderà dal tempo. Inoltre, per un sistema di particelle nello spazio «vuoto», V deve dipendere soltanto dalle posizioni reciproche delle particelle, cioè dalle differenze delle loro coordinate spaziali. La presenza del sistema II può turbare questa simmetria rispetto a tutte le traslazioni spaziali, e V può immediatamente cominciare a dipendere da alcune coordinate o dalle loro funzioni.

Leggi di conservazione e di simmetria. Caratteristiche di estrema importanza di un sistema chiuso sono i suoi integrali di moto: tali funzioni $f(q_i, \dot{q}_i)$

(ed eventualmente derivate di ordine superiore, se da esse dipende la lagrangiana) sono costanti, cioè non dipendono dal tempo, su ogni traiettoria del sistema. Vengono anche dette leggi di conservazione qualora possiedano la proprietà di additività, cioè si sommino in caso di unione di due sistemi senza interazioni (ad esempio, diffusi spazialmente). Si usa questo termine come equivalente a quello di integrale di moto.

Le leggi fondamentali di conservazione sono collegate alle simmetrie del sistema (teorema di Noether). Si supponga che sul fibrato $N \times \mathbf{R}$ agisca un gruppo a un parametro di trasformazioni che conservano la lagrangiana, e che in coordinate locali quest'azione trasformi il punto (t, q_i) nel punto $(f(t, \lambda); f_i(q_j, t, \lambda))$, dove λ è il parametro del gruppo. In questo caso la grandezza

$$\Lambda = \left(L - \sum_i \frac{\partial L}{\partial \dot{q}_i} \dot{q}_i\right) \frac{\partial f}{\partial \lambda} + \sum_i \frac{\partial L}{\partial \dot{q}_i} \frac{\partial f_i}{\partial \lambda}$$

resta costante su qualsiasi traiettoria.

Un sistema meccanico chiuso possiede un gruppo di simmetria a sette parametri: il gruppo a quattro parametri delle traslazioni dello spazio e del tempo, e il gruppo a tre parametri delle rotazioni dello spazio euclideo. Con riferimento a ciò, per un tale sistema esistono sette leggi fondamentali di conservazione: dell'energia, delle tre componenti dell'impulso e delle tre componenti del momento dell'impulso.

La conservazione dell'energia corrisponde al gruppo $t \to t + \lambda$; l'energia è espressa dalla formula

$$E = \sum_i \frac{\partial L}{\partial \dot{q}_i} \dot{q}_i - L.$$

La j-esima componente dell'impulso di un sistema di punti in coordinate cartesiane è

$$\sum_i \frac{\partial L}{\partial x_{ij}} = p_j.$$

Il vettore del momento dell'impulso di questo sistema in coordinate cartesiane è $\sum_i [r_i \cdot p_i]$, dove $p_i = m_i(\dot{x}_{i_1}, \dot{x}_{i_2}, \dot{x}_{i_3})$, $r_i = (x_{i_1}, x_{i_2}, x_{i_3})$ e le parentesi quadre rappresentano il prodotto vettoriale.

La teoria della relatività speciale. La lagrangiana del sistema dei punti materiali

$$L = \sum_{i=1}^{N} \frac{m_i v_i^2}{2} - V(q_{ij})$$

sottintende ancora un'ipotesi di carattere fisico: l'istantaneità della propagazione dell'influenza di una particella sulle altre. Un cambiamento della posizione $(q_{i_1}, q_{i_2}, q_{i_3})$ dell'i-esimo punto si manifesta immediatamente sull'energia potenziale dell'interazione di tutti i punti. Quest'idea dell'*actio in distans* era sembrata insoddisfacente, come argomentazione filosofica, già allo stesso Newton.

Il rigetto totale di quest'idea è alla base dei modelli della teoria della relatività speciale, dove lo spazio-tempo a quattro dimensioni di Minkowski è il veicolo degli eventi. Esso è rappresentato matematicamente come lo spazio affine $\mathbf{R}^4$, munito di sistemi inerziali di coordinate e di una metrica arbitraria ma privilegiata. Se in un dato sistema inerziale (x_1, x_2, x_3, t) e (x_1', x_2', x_3', t') sono le coordinate di due punti P, P', allora il quadrato della distanza a quattro dimensioni fra di essi

$$\sum_{i=1}^{3} (x_i - x_i')^2 - c^2(t - t')^2$$

(c è la velocità della luce) rimane invariato in qualunque altro sistema inerziale.

Da ogni punto P dello spazio-tempo ha origine il cono-luce C_P definito in modo invariante e formato da tutti i punti Q, la cui distanza da P è uguale a zero. Le sue generatrici sono linee universali di particelle che si allontanano da P in modo rettilineo con la velocità della luce. Le particelle (materiali), cioè le particelle di massa m diversa da zero, possono muoversi soltanto con una velocità rigorosamente minore della velocità della luce: la tangente alla traiettoria di una tale particella è situata sempre all'interno del cono-luce che ha origine nel punto di contatto. Perciò l'insieme delle traiettorie virtuali di tale particella è definito come l'insieme (delle parti) delle curve differenziabili nello spazio di Minkowski, le cui tangenti in ogni punto si trovano all'interno del cono-luce. L'azione di una particella libera in una traiettoria virtuale è la sua lunghezza (nel senso della metrica di Minkowski) moltiplicata per la massa col segno meno.

La lagrangiana di un sistema chiuso è invariante rispetto al gruppo di Lorentz delle trasformazioni di un sistema inerziale di coordinate in un altro. Questo gruppo ha dieci parametri, e pertanto il teorema di Noether permette di costruire dieci leggi fondamentali di conservazione per tale sistema: le quattro componenti dell'impulso a quattro dimensioni e le sei componenti del tensore antisimmetrico del momento a quattro dimensioni.

Campo. Uno dei postulati fondamentali della teoria della relatività afferma che nessuna influenza (segnale) può propagarsi con velocità superiore alla velocità della luce. Per questo la descrizione di un sistema di punti interagenti non può essere realizzata mediante il termine V nella lagrangiana del tipo descritto in precedenza. L'analisi conduce a un punto di vista ancora piú radicale in conformità al quale un sistema chiuso di punti interagenti non può in generale essere descritto in termini di uno spazio di dimensione finita nelle loro coordinate: è necessario introdurre lo spazio di dimensione infinita degli stati del campo che descrive l'interazione. Nella maggior parte dei modelli l'interazione è locale, cioè è definita in modo che le particelle eccitino il campo, che il campo agisca su di esse vicino ad esse e che punti vicini del campo interagiscano l'uno con l'altro.

Il campo classico libero, cioè il campo fuori delle particelle, è di solito realizzato matematicamente come sezione di un fibrato vettoriale sullo spazio-tempo M oppure come connessione in questo fibrato.

Ad esempio, il campo elettromagnetico «è» la 2-forma esterna chiusa F su M, oppure il tensore antisimmetrico di rango 2 (è possibile non indicare il carattere della sua varianza in virtú della procedura di innalzamento e di abbassamento degli indici con l'aiuto della metrica). Se v è la velocità a quattro dimensioni di una particella con carica e e con massa m, allora l'azione del campo F su questa particella è descritta dalla sua equazione di moto $dv/d\tau = (e/m)F(v)$. Qui τ è il tempo proprio della particella oppure, il che è lo stesso, la lunghezza della sua traiettoria nella metrica di Minkowski, e F è considerato come un operatore sullo spazio tangente (dove si trova v) con valori nello stesso spazio tangente (dove si trova $dv/d\tau$). Nel sistema inerziale di coordinate $(t, x, y, z) = (x^0, x^1, x^2, x^3)$ le componenti del tensore $F = \sum F_{\alpha\beta} dx^\alpha \wedge dx^\beta/2$ hanno la forma

$$\|F_{\alpha\beta}\| = \begin{Vmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{Vmatrix}$$

dove E, B sono le componenti classiche dell'intensità del campo elettrico e magnetico.

Non si descrivono qui le rimanenti equazioni di Maxwell per non entrare in dettagli tecnici, e si omette anche la descrizione della lagrangiana del campo.

Causalità e analiticità. L'idea di causalità consiste nel fatto che la trasmissione dell'interazione da una parte del sistema a un'altra avviene con ritardo nel tempo. Anche nei modelli di azione a distanza questo ritardo è assunto uguale a zero, ma mai negativo: la causa non può precedere l'effetto. Qui interesserà un aspetto delle descrizioni matematiche di causalità: la comparsa, nei modelli corrispondenti, delle funzioni analitiche di variabile complessa. Le funzioni analitiche hanno la seguente proprietà molto significativa di «rigidezza»: assegnare una funzione in un dominio arbitrariamente piccolo significa definirla in modo completo, cioè assegnarne sia l'intero dominio di definizione che il valore in ogni suo punto. Si può dire che questa è l'espressione estrema dell'idea di dipendenza del comportamento globale dell'oggetto matematico dal suo comportamento locale.

Il meccanismo dell'introduzione delle funzioni analitiche è illustrato da un semplice e classico esempio.

Si supponga di avere un sistema chiuso del tipo di un oscillatore lineare e lo si sottoponga a un «influsso esterno» $f(t)$ dipendente dal tempo t interessandosi poi della «risposta» $g(t)$. Per l'oscillatore lineare, $f(t)$ può presentarsi come una forza che «mette in moto» l'oscillatore, e $g(t)$ come il movimento

risultante. Considerando il sistema come lineare, si può descrivere la risposta come il risultato dell'applicazione a $f(t)$ di un operatore integrale, eventualmente con nucleo di distribuzione: $g(t) = \int_{-\infty}^{+\infty} L(t, t') f(t') dt'$. Se la struttura del sistema non dipende dal tempo, allora $L(t, t')$ deve dipendere soltanto dalla differenza $t - t'$, e la si può scrivere come $L(t - t')$. La condizione di causalità, in particolare, significa che se l'influsso $f(t) = \delta(t - t_0)$ è un impulso dato nell'istante t_0, allora $g(t) = 0$ per $t < t_0$, e quindi $L(\tau) = 0$ per $\tau < 0$. Questo stesso sistema lo si esamini ora in una «descrizione in termini di onde», cioè lavorando, invece che con $f(t)$ e $g(t)$, con le loro trasformate di Fourier

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{+i\omega t} dt$$

e $\hat{g}(\omega)$, $\hat{L}(\omega)$ definite in modo analogo. Dalle proprietà della trasformazione di Fourier che, in particolare, trasforma la convoluzione delle funzioni in t nella moltiplicazione delle funzioni in ω, risulta che $\hat{g}(\omega) = \hat{L}(\omega)\hat{f}(\omega)$. Si ottiene cosí una descrizione molto semplice della risposta in termini dell'influsso. La condizione di causalità si manifesta nel comportamento della funzione $\hat{L}(\omega)$. Precisamente, dal fatto che $L(\tau) = 0$ per $\tau < 0$, segue che $\hat{L}(\omega)$ può essere prolungata come funzione analitica nel semipiano superiore della variabile *complessa* ω ($Im\omega > 0$), oppure che il coefficiente $e^{i\omega\tau}$ per $\tau \geq 0$ in modulo non supera 1 (e decresce rapidamente rispetto a $Im\omega$). La stessa funzione $\hat{L}(\omega)$ per $\omega \in \mathbf{R}$ è il valore al bordo di questa funzione analitica. Oltre a ciò, la condizione $L(\tau) = 0$ per $\tau < 0$ con alcune ipotesi supplementari, che qui non verranno formulate, è equivalente alla seguente relazione tra la parte reale e il coefficiente della parte immaginaria di $\hat{L}(\omega)$:

$$Re\,\hat{L}(\omega) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{Im\,\hat{L}(\omega')}{\omega' - \omega} d\omega',$$

$$Im\,\hat{L}(\omega) = -\frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{Re\,\hat{L}(\omega)}{\omega' - \omega} d\omega'.$$

Qui gl'integrali sono intesi nel senso di valori principali. Queste formule sono dette relazioni di dispersione.

Il loro significato fisico è il seguente. Il passaggio alla trasformata di Fourier significa che si considera il sistema come un filtro, e che si è interessati alla sua risposta alle oscillazioni armoniche che si presentano come una successione infinita stazionaria di onde. Sembrerebbe che il sistema, che lascia passare bene una parte limitata dello spettro e assorbe quasi tutta la parte restante, debba dare in uscita oscillazioni, simili alle armoniche, indipendentemente dal tipo d'ingresso, e perciò «cominciare a risonare» prima che su esso abbia luogo l'influsso. Tuttavia le relazioni di dispersione escludono questa possibilità: la parte reale e il coefficiente della parte immaginaria di $\hat{L}(\omega)$ sono legati in modo che, qualunque sia la parte assorbita dello spettro, il resto ha esattamente quegli spostamenti di fase per cui la risposta non può precedere l'influsso.

Le relazioni di dispersione hanno delle varianti e delle generalizzazioni molto importanti, in particolare nella teoria quantistica dei campi.

Nella relazione iniziale per la comparsa dell'analiticità, insieme alla condizione $L(\tau)=0$ per $\tau<0$, vi è la condizione d'indipendenza del comportamento locale del campo in due punti, divisi da un intervallo di tipo spaziale nell'universo di Minkowski. Essa è espressa nella forma di commutatività degli operatori del campo in questa coppia di punti: siccome un segnale da un punto non può raggiungerne un altro, non mette in comunicazione nessuno di essi.

Si osservi che lo stesso meccanismo dell'analiticità collega, nella meccanica quantistica, anche grandezze di diversa natura.

L'universo come sistema chiuso. I moderni modelli cosmologici sono varietà a quattro dimensioni (spazio-tempo) con una metrica pseudoriemanniana

$$\sum_{\alpha,\beta=0}^{3} g_{\alpha\beta}\,dx^{\alpha}dx^{\beta}$$

localmente simile alla metrica di Minkowski in coordinate localmente inerziali. Il moto della materia in tale spazio è definito dalla sua geometria. Ad esempio, ammettiamo di osservare il movimento di due particelle di prova, vicine nel sistema inerziale locale, con velocità u^{α}; η^{α} sono le componenti della distanza tra di esse. Le particelle si muovono lungo una geodetica, e le (η^{α}) soddisfano il sistema delle equazioni della «deviazione geodetica»

$$\frac{D^2\eta^{\alpha}}{D\tau^2}+\sum_{\beta,\gamma,\delta} R^{\alpha}_{\beta\gamma\delta}u^{\beta}\eta^{\gamma}u^{\delta}=0.$$

Qui τ è il tempo proprio del sistema, $D/D\tau$ è la cosiddetta «derivata assoluta» lungo di esso, $R^{\alpha}_{\beta\gamma\delta}$ è il tensore della curvatura dello spazio espresso mediante le componenti della metrica $g_{\alpha\beta}$.

A sua volta, la curvatura dello spazio-tempo deriva dalla materia presente in esso in virtú delle equazioni di Einstein

$$R_{\alpha\beta}-\frac{1}{2}g_{\alpha\beta}R=8\pi T_{\alpha\beta}.$$

Qui $R_{\alpha\beta}$, R, sono le contrazioni di $R^{\alpha}_{\beta\gamma\delta}$ rispetto agli indici corrispondenti, e $T_{\alpha\beta}$ è il tensore impulso-energetico della materia (compreso il campo) che è costruito mediante una formula nota.

Le equazioni di Einstein hanno un gran numero di soluzioni, e la scelta dei candidati per i modelli cosmologici fra di esse è effettuata tenendo conto dei dati astronomici osservati. Nel piú semplice modello attualmente ammesso si suppone che la materia nell'Universo sia un mezzo omogeneo, continuo e isotropo. Le non-omogeneità osservate nelle proporzioni $\leq 10^8$ di anni luce si possono considerare come perturbazioni su una base omogenea in media. In tale modello esiste un sistema naturale di coordinate: gli assi spaziali delle coordinate sono le linee universali dei «punti» della materia, l'asse temporale è il tempo proprio di ogni linea universale, uguale a zero su una fissata iper-

superficie omogenea di tipo spaziale. Supponendo che la densità $\rho(t)$ e la pressione $p(t)$ della materia siano conosciute, allora la geometria del modello è una funzione del tempo $a(t)$, il «raggio dello spazio». In conseguenza delle equazioni di Einstein tali funzioni verificano:

$$\left(\frac{\dot a}{a}\right)^2-\frac{8}{3}\pi\rho=-\frac{k}{a^2},$$

$$\dot\rho+3(p+\rho)\frac{\dot a}{a}=0.$$

Il parametro k nelle unità stabilite può essere assunto uguale a ± 1 oppure a 0. Per quel che riguarda p e ρ, sono possibili diverse ipotesi. I seguenti due modelli conducono facilmente alle equazioni integrabili degli «universi di Friedman»: $p=0$, $\rho\sim a^{-3}$ («universo corpuscolare»); $p=\rho/3$, $\rho\sim a^{-4}$ («universo radiante»). Il caso $k=+1$ conduce al modello «chiuso» in tutti i tempi dell'universo, i casi $k=0$ e $k=-1$ conducono al modello «aperto» (rispettivamente, piano o iperbolico).

Senza entrare in ulteriori particolari (la descrizione di altri modelli, i dati sperimentali e la loro interpretazione, le possibilità di altre leggi di evoluzione del mondo, ecc.), si studieranno qui brevemente le questioni filosofiche poste dal considerare l'universo come sistema isolato. Esse in ultima analisi sono due: il problema dell'unicità dell'universo e il problema delle condizioni iniziali.

Come si è già detto, i principî fondamentali della descrizione dei sistemi isolati sono basati sull'ipotesi della loro ripetuta riproducibilità: lo spazio degli stati e la descrizione nello spazio delle fasi del sistema realizzano l'idea che «esemplari diversi» del sistema possono attraversare diverse possibili tappe dell'evoluzione. Com'è possibile rendere compatibili questi principî con l'unica rappresentazione dell'universo accessibile alla nostra osservazione? Poiché i nostri modelli sono basati sull'idea dell'interazione locale, l'utilità della deduzione di queste o quelle equazioni dell'evoluzione consiste nella verifica della loro validità locale nello spazio-tempo, al variare della distribuzione della materia.

Il problema delle condizioni iniziali è piú complesso per diverse ragioni. Non si sa da che cosa siano determinate le condizioni iniziali della fase osservata dell'evoluzione dell'universo la quale, evidentemente, era una fase superdensa. Bisogna anche tener presente che, nei primissimi stadi di compressione, il modello classico diventa troppo sommario e cominciano ad avere un ruolo dominante gli effetti quantistici, ma non esiste un'unica teoria della gravitazione quantistica. Anche nei limiti del modello classico rimane inesplicabile il problema delle cause dell'omogeneità osservata su larga scala. Quest'ultima presuppone un influsso reciproco delle parti del mondo talmente forte, da condurre a un livellamento delle caratteristiche. Ma ad ogni punto universale possono essere causalmente collegati solo i punti universali che si trovano all'interno o sulla superficie del suo cono-luce (nello spazio curvo): all'interno cioè dell'«orizzonte» di tale punto. Nel mondo di Friedman, per estrapolazione all'indietro, la massa della materia che si trova all'interno dell'oriz-

zonte sembra tendere a zero al momento della compressione. In altre parole l'universo, nei primi momenti, si frantuma in molte regioni non collegate causalmente, e non c'è motivo che le caratteristiche della materia coincidano in tali regioni. I tentativi di venire a capo di questa difficoltà hanno condotto Meissner a considerare modelli anisotropi, nei quali l'universo è piú simile a una ellissoide con tre semiassi diversi, che non a una sfera. L'evoluzione di questi modelli nel tempo passato rispetto al momento di espansione è di tipo oscillatorio. In questo regime si suppone che le oscillazioni dell'anisotropia permettano, nelle prime fasi, il diffondersi per tutto l'universo della radiazione, provocando in tal modo il suo «rimescolamento» («mixmaster universe»). La materia potrebbe in tal modo diventare omogenea, le oscillazioni si attenuerebbero con l'andar del tempo e il modello si avvicinerebbe al modello dell'universo di Friedman. Tuttavia, calcoli dettagliati mostrano che, per la maggior parte delle condizioni iniziali, il tempo realmente passato dall'inizio dell'espansione dell'universo risulta insufficiente per la sua riduzione allo stato omogeneo, e il problema rimane insoluto.

L'aspetto della cosmologia che qui interessa, cioè il problema della reciproca dipendenza delle parti dell'universo, si presenta molto chiaramente ancora in una serie di modelli fisici, in questo caso di carattere piú locale, cioè nei modelli dei «buchi neri».

Buchi neri. Secondo le equazioni di Einstein, la massa-energia incurva lo spazio. Questa curvatura può essere cosí grande da chiudere una qualche regione dello spazio nel senso che qualunque curva nello spazio-tempo, fatta uscire dalla sua frontiera, dovrà avere delle parti che si trovano fuori dal conoluce, e perciò non realizzate come linee universali di particelle di massa non nulla oppure di raggi di luce. Tale regione è detta buco nero [cfr. Rees, Ruffini e Wheeler 1975]. Nessun evento accaduto all'interno di essa può, in quanto causa, influenzare gli eventi al suo esterno; la radiazione che cade fuori di essa viene assorbita. Il processo di formazione del buco nero non può consistere nel collasso gravitazionale, cioè nella compressione di una nube di materia fino ad una densità con la quale la curvatura dello spazio diventa sufficientemente grande per la sua chiusura.

3.3. Dipendenze quantistiche.

I tratti essenziali della diversità delle descrizioni classiche e quantistiche possono essere riassunti nel modo seguente.

Lo spazio degli stati del sistema quantistico si presenta sempre come uno spazio vettoriale complesso $\mathcal{H}$ con una distanza, come uno spazio di Hilbert. Con maggior precisione, è lo spazio dei raggi $\{\lambda\psi\}$, $\lambda \in \mathbf{C}\,(|\lambda| = 1)$, $\psi \in \mathcal{H}$: gli stati descritti dai vettori ψ e $\lambda\psi$ sono fisicamente identici. Perciò vale il principio di sovrapposizione: se il sistema può trovarsi nelle posizioni ψ_1 e ψ_2, allora può anche trovarsi in qualsiasi posizione del tipo $\lambda\psi_1 + \mu\psi_2$, $\lambda, \mu \in \mathbf{C}$. Nella fisica classica non c'è niente di simile: una combinazione lineare formale

delle posizioni della pallina, che si trova nel punto q_1 e nel punto q_2, non corrisponde alla realtà macrofisica. (Le note eccezioni quantistiche a questo principio sono le regole di superselezione che vietano le combinazioni lineari degli stati con differenti numeri quantici, ad esempio, della carica o del numero barionico, ma qui non ci si addentrerà in ulteriori particolari). La possibilità della superposizione conduce a una relazione molto particolare di «interdipendenza» degli stati distinti dell'uno o dell'altro sistema, col risultato che l'osservazione macroscopica del sistema in qualunque stato permette di determinare con ben definita probabilità il valore delle sue caratteristiche tipiche per qualunque stato.

Parlando piú precisamente, siano A l'operatore dell'osservabile quanto-meccanico e ψ_1, ψ_2 gli stati nei quali l'osservabile ammette rispettivamente valori esatti λ_1 e λ_2. Allora, misurando A per il sistema nello stato $\psi = a_1\psi_1 + a_2\psi_2$, si ottiene il valore λ_i con probabilità $|a_i|^2/\|\psi\|^2$, dopo che il sistema si è portato nello stato ψ_i (se è non-singolare). Perciò il processo di misurazione del sistema in un determinato momento è il processo della sua preparazione in un altro stato. Nella meccanica classica la misurazione non influisce sullo stato del sistema.

Questo modello di misurazione non dev'essere primario nella successiva teoria. Dev'essere il riflesso del fatto che si osserva in modo classico lo strumento di misura, e in modo quanto-meccanico il sistema misurato e, non possedendo informazioni dettagliate sulla posizione quantistica dello strumento, si effettua qualche media lungo il suo spazio degli stati, il quale influenza l'introduzione sia dell'operatore A sia delle previsioni di probabilità. Nel processo di misurazione si unificano due sistemi, e nella teoria completa si dovrebbe seguire il comportamento del vettore degli stati del sistema unificato. Tuttavia ciò è possibile soltanto con l'aiuto dell'aggiunta di un terzo macrostrumento, necessario per la misurazione del sistema unificato, e qui di nuovo sorgono le difficoltà. Hermann Weyl ha confrontato questa situazione con la necessità, dimostrata da Gödel, di una gerarchia di linguaggi formali di potenza sempre crescente per approssimare l'idea di verità matematica.

Se $\mathcal{H}_1$, $\mathcal{H}_2$ sono gli spazi degli stati di due sistemi quantistici I e II, allora lo spazio degli stati della loro unione è contenuto in, ma non necessariamente coincide con, $\mathcal{H}_1 \otimes \mathcal{H}_2$. In $\mathcal{H}_1 \otimes \mathcal{H}_2$ ci sono dei vettori «scomponibili» $\psi_1 \otimes \psi_2$, con $\psi_i \in \mathcal{H}_i$, che possono presentarsi intuitivamente come stati nei quali «il sistema I si trova nella posizione ψ_1, e il sistema II nella posizione ψ_2». Ma tutto $\mathcal{H}_1 \otimes \mathcal{H}_2$ è ben lungi dall'essere formato solo da vettori scomponibili. Secondo il principio di sovrapposizione, sono ammissibili anche gli stati del tipo $\psi_1 \otimes \psi_2 + \psi_1' \otimes \psi_2'$, e intuitivamente ciò significa che tale stato del sistema «I+II» non può essere descritto assegnando «separatamente» gli stati del sistema I e II. Perciò già lo stesso principio di sovrapposizione impedisce di osservare I e II come «parti integranti» di I+II, almeno fino al momento in cui I e II non interagiranno fisicamente. Dopo la fine dell'interazione («l'elettrone ha attraversato il settore del campo magnetico», ecc.) diventa di nuovo possibile in modo approssimativo l'osservazione separata dei sistemi anche se non in tutte le situazioni, come mostra il seguente famoso esperimento immaginario.

Esperimento di Einstein-Rosen-Podol'skij. Lo si esporrà nella variante di Bohm. Nell'istante iniziale si abbia una molecola con due atomi in uno stato con spin nullo; lo spin di ogni atomo sia uguale a $\hbar/2$, ed essi siano orientati in modo opportuno. Disintegrata questa molecola, senza cambiare il momento rotatorio, si aspetta fino a che gli atomi separati non cessino d'interagire in modo evidente. S'immagini di misurare lo spin di uno degli atomi. Lo spin si presenta come una grandezza vettoriale, e gli operatori di proiezione dello spin σ_x, σ_y, σ_z su ognuno degli assi non commutano. Perciò in un solo esperimento è possibile misurare soltanto una delle componenti, ad esempio σ_z. Viene pronosticata soltanto la ripartizione delle probabilità del risultato della misura. In virtú della legge di conservazione del momento, la misura attuale di σ_z per uno degli atomi conduce a predire che il valore di σ_z per l'altro atomo nello stesso momento è uguale in modulo ed è opposto in direzione, benché non vi siano interazioni effettive tra gli atomi. Per di piú, non eseguendo la misurazione di σ_z per l'atomo *I*, ma soltanto immaginandola per i diversi orientamenti dell'asse *z*, si potrebbe concludere che l'atomo II «possiede» precisi, ancorché ignoti, valori della proiezione dello spin in tutte le direzioni. Essi sono precisi perché le misurazioni per l'atomo I non perturbano l'atomo II. Ciò contraddice l'ipotesi secondo cui la funzione ψ dell'atomo II contiene tutta l'informazione a proposito di esso, e l'esperimento di Einstein-Rosen-Podol'skij era stato ideato anche per dimostrare la completa insufficienza della descrizione quantistica. Una precisa analisi quanto-meccanica, tuttavia, dimostra che la contraddizione è presente altrove: nell'ipotesi secondo cui l'atomo II può essere osservato come un sottosistema isolato del sistema I+II, benché tra I e II non vi sia apprezzabile interazione «di forze». In realtà, le caratteristiche degli atomi I e II sono correlate per mezzo del principio di sovrapposizione, che non ammette una reinterpretazione classica soddisfacente.

L'origine dell'universo da una piccola regione di materia superdensa conduce probabilmente alla completa interazione quantistica di tutte le sue parti, cioè all'inatteso ritorno a un'antica idea della filosofia orientale.

Particelle identiche: statistiche di Fermi-Dirac e Bose-Einstein. Sia $\mathcal{H}$ lo spazio degli stati di una particella. Allora lo spazio degli stati $\mathcal{H}^{(n)}$ di un sistema di *n* tali particelle è contenuto in $\mathcal{H}^{\otimes n}$, prodotto tensoriale *n*-esimo dello spazio $\mathcal{H}$. Tuttavia quello non coincide affatto con questo in virtú del principio dell'indistinguibilità delle particelle: qualunque permutazione delle *n* particelle non cambia lo stato del loro sistema. Perciò $\mathcal{H}^{(n)} \subset \mathcal{H}^{\otimes n}$ si compone sia della parte simmetrica (statistica di Bose-Einstein), sia della parte antisimmetrica (statistica di Fermi-Dirac), e in $\mathcal{H}^{(n)}$ non esistono in generale stati che potrebbero essere descritti come l'insieme «*j*-esimo stato della *i*-esima particella», cioè sono tutti quanti rimescolati dalle trasformazioni di simmetria. Da qui deriva l'importante principio di Pauli per gli elettroni dove $\mathcal{H}^{(n)} = \Lambda^n \mathcal{H}$ (prodotto antisimmetrico): nessuna coppia di elettroni in un atomo può trovarsi in uno stesso stato. (Va qui sottolineata un'ulteriore inadeguatezza della lingua naturale per la descrizione del formalismo matematico: si è appena osservato che gli elettroni sono indistinguibili, ma si parla di elettroni diversi). Il tutto funziona come se i fermioni si «respingessero» l'un l'altro, benché qui non sia presente nessuna forza di repulsione (la repulsione elettromagnetica delle particelle di ugual carica non è in rapporto con questo fatto: la statistica è legata allo spin e non alla carica). Al contrario, i bosoni «cercano di ammassarsi» nella stessa posizione. I fotoni si presentano come bosoni, e questo meccanismo si trova nell'operazione essenziale del laser di generare una radiazione coerente di grande potenza. Senza dubbio qui non sono presenti nemmeno le classiche «forze di attrazione». La specifica interazione quantistica nei sistemi di particelle identiche è descritta in termini di simmetria e di superposizione delle loro funzioni d'onda ψ.

Interferenza delle alternative classiche e integrali di Feynmann. Il principio di sovrapposizione riceve un'originale interpretazione nella riformulazione di Feynmann per la meccanica quantistica [cfr. Feynmann e Hibbs 1965]. Si consideri un sistema classico che sia descritto dall'insieme delle traiettorie virtuali *E* e dal funzionale dell'azione $S : E \to R$. Questo modello è la «materia prima» per la successiva descrizione quantistica. Nella teoria quantistica il problema non consiste nella separazione classica delle traiettorie ammissibili del sistema da quelle virtuali – queste traiettorie classiche non hanno un senso proprio nel modello dato, ma nel calcolo dell'ampiezza (a valori complessi) di probabilità della transizione da ben definiti stati iniziali a quelli finali. La stessa probabilità di transizione è definita allora come il quadrato del modulo dell'ampiezza di probabilità. Secondo la concezione di Feynmann, per il calcolo dell'ampiezza di probabilità bisogna considerare su *E* il funzionale e^{iS} (qui *S* è misurata nelle unità d'azione di Planck), e integrarlo in un certo senso su quella parte delle traiettorie virtuali di *E* che corrispondono al passaggio da un dato stato iniziale ad un dato stato finale. Questa ampiezza, come funzione dello stato finale rispetto a uno iniziale non definito, è la funzione d'onda ψ.

Si osservi che i concetti di stato iniziale e finale, a differenza della traiettoria tra di essi, conservano un senso semiclassico poiché in sostanza si riferiscono attualmente almeno a due processi sperimentali immaginari: l'«allestimento» del sistema quantistico nella posizione iniziale e l'atto di «misurazione» in quella finale. S'è già detto che queste procedure consistono nell'interazione del sistema quantistico con strumenti macroscopici, che sono descritti in modo classico.

Quindi, nello schema di Feynmann, tutte le storie virtuali classiche del sistema portano il loro contributo, oltre a ciò nel tipo di struttura del funzionale e^{iS} questi contributi interferiscono; ciò conduce anche alla descrizione ondulatoria tipica per la meccanica quantistica, ad esempio alla diffrazione degli elettroni su un reticolo cristallino. La conservazione degli effetti di diffrazione con intensità molto piccole dimostra che le traiettorie classiche possibili di un elettrone interferiscono per l'appunto tra di loro, ma ciò non accade per le traiettorie classiche effettive di alcuni elettroni, cioè l'ultimo concetto non è adeguato alla realtà fisica. Questo è uno degli sviluppi sorprendenti della di-

pendenza di tipo non classico che fornisce una buona illustrazione del fatto che la dipendenza appare come termine metalinguistico, entrato piuttosto nella descrizione del modello matematico, che non in quella della modellizzazione della sua realtà. L'effetto d'interferenza è legato al contributo delle traiettorie virtuali alla fase, nell'ampiezza di probabilità, ma queste fasi non si presentano ancora come grandezze fisiche osservate perché non si riferiscono al processo reale.

3.4. Interazioni termodinamiche.

La termodinamica ha a che fare con i sistemi formati da un numero molto grande di particelle, come in un volume di gas, e studia le loro caratteristiche macroscopiche. Queste caratteristiche in ultima analisi sono soggette a interpretazione come medie statistiche nello spazio degli stati microscopici del sistema, e il loro comportamento macroscopico dev'essere interpretato nei modelli statistici tipo quelli considerati nel § 2. Ciò è legato a un'intera cerchia di difficili e non completamente risolti problemi matematici, e qui ci si limiterà alla descrizione qualitativa di una serie di concetti fondamentali.

Lo spazio degli stati di un sistema grande. Nella variante classica come in quella quantistica un modello per lo spazio degli stati è la terna formata dall'insieme degli stati M, dalla funzione $E : M \to \mathbf{R}$ ($E(x)$ è l'energia del sistema nella posizione x) e dalla misura μ su M. Nel caso di un sistema classico hamiltoniano, in qualità di M si assume lo spazio delle fasi, in qualità di μ la misura di Liouville, che è invariante rispetto al flusso del movimento lungo le traiettorie nello spazio delle fasi. Nel caso di un sistema quantistico, M è spesso ritenuto finito benché sia un sistema molto grande, e la misura è la probabilità che il sistema si trovi in un dato sottoinsieme di stati.

Tutta la terna (M, E, μ) può dipendere da alcuni parametri microscopici: le misure del serbatoio contenente il gas, il campo magnetico esterno nel quale esso si trova, ecc. Questi parametri sono detti esterni. Assegnando i parametri esterni e un piccolo intervallo energetico $(E, E+\delta E)$ si determina un insieme di (micro) stati accessibili per il sistema con questi parametri. Su questo insieme μ induce una nuova misura, la probabilità.

Il sistema è detto in equilibrio se la probabilità di trovarsi in una qualunque regione di microstati accessibili non dipende dal tempo.

Va sottolineata la differenza dal sistema formato dalla pallina nella scanalatura: le sue posizioni di equilibrio sono le traiettorie nelle quali essa è immobile, cioè non è mutata nessuna delle sue caratteristiche. Gli stati di equilibrio di un sistema grande sono gli stati nel quale sono costanti soltanto le medie delle caratteristiche degli stati accessibili. Il gas con una certa pressione, volume e temperatura si trova in equilibrio nonostante i complessi movimenti microscopici compiuti dalle molecole isolate.

Entropia e temperatura. Si indichi con $\varphi(E)$ la misura del sottospazio degli stati con energia $\leq E$. La funzione $S(E) = \log(d\varphi(E)/dE)$ è detta entropia, e la funzione $(dS/dE)^{-1}$ è detta temperatura formale del sistema in equilibrio con energia E. L'entropia, come ordine di grandezza, è proporzionale al numero dei gradi di libertà del sistema, che è molto grande (si ricordi che il numero di Avogadro è dell'ordine di 10^{23}).

Interazione termica. Siano I e II due grossi sistemi e li si riunisca in un unico sistema con energia totale E^*. Si suppone che l'interazione fra le parti I e II del sistema unificato abbia luogo grazie allo scambio stocastico delle energie in modo che questi atti di scambio quasi non turbino l'evoluzione indipendente dei sottosistemi I e II nel loro spazio degli stati accessibili. Ciò significa che la densità di probabilità, tale che il sottosistema I abbia l'energia E ed il sottosistema II l'energia E^*-E, è proporzionale a $d\varphi_1(E)/dE \times \times d\varphi_2(E^*-E)/dE$. Ma con l'aumentare di E il primo fattore cresce molto in fretta, e il secondo si abbassa molto in fretta. Quindi questa densità di probabilità ha un picco molto limitato che corrisponde a un estremo della derivata di questa funzione di E, cioè alla condizione di uguaglianza delle temperature formali dei sottosistemi I e II. Dal momento che $\varphi(E)$ cresce con l'aumentare di E, insieme alla sua derivata, come una potenza molto alta di E, la temperatura formale del sistema è sempre non-negativa. Con l'interazione termica dei sistemi, le loro temperature in ultima analisi si livellano; l'energia passa dal sistema con una temperatura maggiore al sistema con una temperatura minore. Lo stato limite ha un'entropia massima. Dal momento che l'entropia è una misura logaritmica del numero degli stati accessibili, e la traiettoria microscopica nello spazio delle fasi del sistema riempie in modo uniforme il corrispondente sottospazio, l'entropia massima risponde al massimo «stato caotico» del microcomportamento del sistema, all'assenza di una strutturazione macroscopica, per quanto ciò sia compatibile con le condizioni esterne.

Quest'idea conduce a una concezione molto importante per la biologia teorica, costringendo a prestare particolare attenzione a come gli organismi viventi si conservano e a come conservano la loro complessissima organizzazione strutturale. Stando a Schroedinger, l'organismo «assorbe un'entropia negativa» dai generi alimentari e fa uso dell'energia esterna per lottare contro i processi termodinamici generali di allineamento termico e di convergenza verso un caos omogeneo. Questo risultato è ottenuto a spese dell'aumento dell'entropia dell'ambiente immediatamente circostante. Negli ultimi dieci anni questo problema ha acquisito un inaspettato valore sociologico nelle questioni di conservazione dell'ambiente ecologico in cui vivono gli uomini. Lo stesso ambiente ecologico naturale è un'entità altamente strutturata e il suo inquinamento, provocato dagli scarti della produzione e della vita quotidiana, è l'immagine visibile del prezzo per la diminuzione dell'entropia, cioè per l'organizzazione della natura sociale tecnologica della società contemporanea.

Insieme canonico. Si consideri il caso dell'interazione termica di due sistemi I e II, dei quali il primo è molto piccolo rispetto al secondo (nel senso del numero di gradi di libertà).

Il sistema II viene detto termostato se è cosí grande che la sua temperatura nell'interazione col sistema I è T. Dall'osservazione precedente non è difficile dedurre che la probabilità di trovare il sottosistema I nello stato di energia E è proporzionale a $e^{-E/T}$. Il fattore di proporzionalità è determinato dalla condizione che la probabilità totale sia uguale all'unità. (Se la temperatura t è misurata in gradi, allora la temperatura T è uguale a kt, dove k è la costante di Boltzmann).

L'insieme dei sistemi con equazioni dell'energia $\{E_r\}$ e con ripartizione delle probabilità del tipo

$$P_r = e^{-E_r/T}\left(\sum_s e^{-E_s/T}\right)^{-1}$$

è detto insieme canonico. È uno dei modelli fondamentali della fisica statistica che rappresenta la classe fondamentale delle interazioni del sistema con l'ambiente circostante.

Si osservi l'analogia formale della «somma statistica» $\sum_s e^{-E_s/T}$ con la «somma secondo la traiettoria virtuale» di Feynmann $\sum_s e^{iS_s/h}$. In entrambi i casi nell'indice degli esponenti si trova una grandezza che misura l'azione, ma nell'integrale di Feynmann c'è inoltre l'unità immaginaria che da una parte determina l'interferenza delle ampiezze, e dall'altra conduce al problema matematico straordinariamente difficile della convergenza. Una delle impostazioni per una rigorosa definizione e per il calcolo degl'integrali di Feynmann consiste nello sfruttare quest'analogia: l'integrale di Feynmann converte la somma statistica mediante il lavoro con «un tempo puramente immaginario». Questa sostituzione trasforma lo spazio-tempo di Minkowski nello spazio-tempo euclideo a quattro dimensioni. L'analogia con la fisica statistica, la tecnica piú semplice di lavoro nella teoria euclidea dei campi quantistici, con il collegamento all'idea di prolungamento analitico, è un metodo molto importante nella fisica teorica degli ultimi anni.

Termodinamica macroscopica e interazioni termodinamiche generali. Nella fisica classica la termodinamica era stata creata come scienza relativa ai parametri macroscopici dei volumi di gas, sottoposti a influenze esterne che forniscono o prelevano calore e lavoro meccanico. Se ne esporranno brevemente i risultati seguendo Truesdell. La giustificazione degli assiomi matematici di partenza e la definizione dei parametri macroscopici come medie possono essere realizzate con strumenti della fisica statistica a partire dagli stati microscopici, nello spirito della precedente esposizione. La differenza del punto di vista qui ammesso è ricca di contenuto nel senso che l'ambiente esterno al sistema termodinamico non è piú considerato come un sistema simile, ma è descritto in termini di trasmissione di calore e di lavoro meccanico.

Lo stato istantaneo del sistema è dato dai quattro numeri (V, T, p, Q) (volume, temperatura, pressione, calore). Il processo è descritto da quattro funzioni $V(t)$, $T(t)$, $p(t)$, $Q(t)$, dipendenti dal tempo $t \in [t_1, t_2]$. Nella descrizione del processo reale esse devono essere considerate come proprietà intrinseche del sistema e anche come caratteristiche della sua interazione con l'ambiente esterno. Gli assiomi della termodinamica postulano le seguenti restrizioni cui devono soddisfare V, T, p, Q durante i processi reali. Esse sono molto interessanti dal punto di vista matematico.

a) *Equazione di stato.* Esiste una funzione continua $\pi(V, T)$, con derivate parziali continue rispetto a V e T, tale che $p = \pi(V, T)$ in un insieme $\mathfrak{D}$ di volumi e di temperature.

b) *Calore latente e calore specifico.* Esistono funzioni continue con derivate parziali continue Λ_v e K_v, tali che per qualunque processo

$$Q = \Lambda_v(V, T)\dot{V} + K_v(V, T)\dot{T}$$

(il punto indica la derivazione rispetto al tempo).

Un processo in cui $Q = 0$ è detto adiabatico, e la curva corrispondente nello spazio delle fasi è detta adiabatica. Il processo in cui $T = \text{cost}$ è detto isotermo, e la curva corrispondente è detta isoterma.

Il lavoro prodotto dal sistema durante il processo $\mathcal{P}$ è uguale a

$$\int_{\mathcal{P}} \pi(V, T)\, dV = L(\mathcal{P}).$$

Le grandezze

$$C(\mathcal{P}) = \int_{t_1}^{t_2} Q\, dt \qquad C^{\pm}(\mathcal{P}) = \frac{1}{2}\int_{t_1}^{t_2} (|Q| \pm Q)\, dt$$

rappresentano rispettivamente l'incremento netto di calore, il calore assorbito dal sistema e il calore ceduto dal sistema durante il processo.

Il processo è detto ciclo di Carnot se è formato da due adiabatiche che uniscono due isoterme, e inoltre $Q > 0$ (il calore è assorbito) sull'isoterma con la temperatura piú alta T^+, e $Q < 0$ (il calore è ceduto) sull'isoterma fredda T^-. È possibile dimostrare che esistono delle funzioni g e h di una variabile, tali che per ogni ciclo di Carnot $\mathcal{C}$ con temperatura superiore T^+ e inferiore T^- risulta

$$\frac{L(\mathcal{C})}{C^+(\mathcal{C})} = \frac{g(T^+) - g(T^-)}{h(T^+)}.$$

La funzione h è definita con esattezza a meno di una costante moltiplicativa, e per una data h la funzione g è definita a meno di una costante additiva. Questo risultato deriva dall'assioma di Carnot secondo il quale $L(\mathcal{C})$ dipende in modo universale soltanto da T^+, T^- e da $C^+(\mathcal{C})$. Allora per qualunque processo ciclico $\mathcal{P}$ abbiamo

$$L(\mathcal{P}) = \int_{\mathcal{P}} \frac{g}{h}(\Lambda_v\, dV + K_v\, dT).$$

Nell'insieme $\mathfrak{D}$ esiste una funzione $E(V, T)$, detta energia interna del sistema, per la quale $\dot{E}=(g/h)Q-p\dot{V}$ in ogni processo. (Essa dipende da g, h in modo non essenziale).

Ammettiamo ancora che nell'insieme $\mathfrak{D}$ esistano una funzione positiva continua f e una funzione differenziabile H tali che $H=Q/f$ per i processi reali. Allora la grandezza $\varphi=E-gH$ è detta potenziale termodinamico del sistema. Assegnandolo si definiscono tre funzioni fondamentali che caratterizzano il processo: l'equazione di stato $\pi(V, T)=p$ e Λ_v, K_v con le formule

$$\pi=-\frac{\partial\varphi}{\partial V}, \qquad \Lambda_v=-\frac{h}{g}\frac{\partial^2\varphi}{\partial V\,\partial T}, \qquad K_v=-h\frac{\partial}{\partial T}\left(\frac{1}{g'}\frac{\partial\varphi}{\partial T}\right).$$

Per i gas ideali si può porre $h=T$, $g=T+\text{cost}$ e $Q=T\dot{H}$ dove H è l'entropia come funzione nell'insieme $\mathfrak{D}$.

3.5. Stabilità, segnale, informazione.

Stabilità. Nella concezione di sistema isolato è sottintesa una certa invarianza delle sue caratteristiche nel dominio spazio-temporale della sua esistenza, che permette soltanto d'identificare questo sistema come tale. In questo senso un'idea primordiale di stabilità è già presente nei concetti scientifici fondamentali. Tuttavia questo termine è usato piú frequentemente in accezioni piú specializzate con riferimento *a*) alle caratteristiche degli uni o degli altri regimi di evoluzione del sistema, *b*) alle caratteristiche dei modelli matematici che descrivono il sistema.

Nell'accezione *a*) si è usata questa parola in connessione con gli stati di equilibrio di una pallina nel fondo di una buca di potenziale. Nell'accezione *b*) il termine è stato usato per la descrizione di quei modelli le cui piccole variazioni conducono a piccole variazioni delle caratteristiche previste del sistema. È evidente il ruolo peculiare del comportamento dei regimi stabili e dei modelli matematici stabili. In particolare la stabilità di un *modello* è importante per essere sicuri di poterlo *applicare*: se una piccola variazione dei dati iniziali conduce alla rapida dispersione delle traiettorie del sistema meccanico, come nei flussi geodetici sulle superfici di curvatura negativa, allora la descrizione dettagliata delle traiettorie individuali perde senso. Nella realtà non possiamo misurare i dati iniziali con precisione arbitrariamente grande; oltre a ciò, le fluttuazioni e le deviazioni dal modello idealizzato agiranno per tutta la lunghezza della traiettoria. Bisogna ancora separare l'insieme delle caratteristiche previste e interessarsi a proprietà statistiche o topologiche delle traiettorie, come in termodinamica.

Bisogna ancora notare l'indeterminatezza del termine 'piccole variazioni del modello'. Spesso viene interpretato come l'introduzione di piccoli parametri in particolari equazioni del modello. Ma gli esempi della teoria della relatività speciale rispetto alla meccanica newtoniana («il piccolo parametro è v/c») e della meccanica quantistica rispetto alla fisica classica («il piccolo parametro è h») dimostrano che i modelli fondamentali di diversi livelli non sono affatto ridotti l'uno all'altro annullando semplicemente il piccolo parametro. Quest'ultimo fatto può essere valido soltanto in una certa regione degli effetti previsti da questi modelli; la struttura generale della teoria risulta assolutamente diversa, e nelle altre regioni gli effetti possono essere assolutamente diversi da quelli previsti.

Instabilità. L'instabilità di alcuni regimi critici di comportamento del sistema è alla base degli effetti di amplificazione tipo «meccanismo di discesa». Soltanto grazie all'esistenza di tali regimi nei macrosistemi gli effetti microscopici possono essere amplificati ed osservati. Un famoso esempio della fisica sperimentale è quello delle tracce delle particelle elementari nella camera di Wilson. Tuttavia, fino ed oltre la nascita di un'esperienza fisica, tutto il comportamento degli organismi viventi è legato a tale instabilità. La vista e l'udito, cioè i canali fondamentali che legano gli organismi superiori al mondo esterno, introducono nell'organismo quantità di energia trascurabili in confronto al dispendio muscolare nel processo dell'attività vitale. Tuttavia sono proprio queste quantità che mettono in moto i meccanismi di caccia del predatore e di fuga della preda. Le loro caratteristiche quantitative non hanno nessun ruolo in confronto a quelle qualitative, come veicoli di segnali o d'informazioni di un determinato tipo. L'organismo vivente che sta all'erta appare dal punto di vista dalla fisica come un sistema instabile di grado superiore; la sua struttura piú complessa assicura la canalizzazione dei vari comportamenti possibili in un andamento biologicamente efficiente, la scelta dei quali viene realizzata in base all'integrazione, continuamente variata, del quadro informativo derivato da deboli influssi, sia esterni che interni.

Nella terminologia di Thom la descrizione dettagliata nello spazio delle fasi dell'organismo vivente deve continuamente subire revisioni catastrofiche, legate alla scomparsa e alla nascita di attrattori, alla modificazione della configurazione delle regioni di attrazione, al passaggio di attrattori instabili alla stabilità e viceversa.

È molto probabile che l'esistenza stessa di una lingua, la sua struttura e la realizzazione delle sue funzioni siano legate alla descrizione delle instabilità catastrofiche, quasi «scheletro discreto» del mondo e della sua storia. L'analisi di Thom [1974], ad esempio, ha mostrato il legame molto stretto che lega la semantica di una grande classe di verbi all'elenco delle «catastrofi elementari» nello spazio-tempo a quattro dimensioni. Risulterebbe forse utile, da questo punto di vista, l'analisi della semantica del linguaggio matematizzato della scienza.

L'osservazione dei modelli scientifici dimostra che nella maggioranza di essi prevale decisamente o l'aspetto continuo della descrizione, legato all'idea di regime stabile, o l'aspetto discreto, legato alla regolarità del passaggio attraverso una catastrofe o all'interdipendenza delle catene di catastrofi.

Estremamente difficile risulta la ricerca di una descrizione sintetica, nella quale il regime catastrofico sia analizzato come caso limite sullo sfondo degli instabili, ma i regimi stabili a un livello piú dettagliato di osservazione si pre-

sentino come medie di quelli catastrofici. Tali sono, in senso lato, i problemi del cambiamento di fase nella fisica statistica, il legame tra la concezione della teoria quantistica dei campi e le matrici di dispersione, e la maggior parte dei problemi della biologia, in particolare della biologia dell'evoluzione.

3.6. Considerazioni su modelli non fisici.

Da quanto detto in precedenza dev'essere chiaro che lo studio della vita, dai virus agli uomini, nella misura in cui può essere fondato sulle idee tradizionali della fisica, deve in un modo o nell'altro essere messo in relazione con il complesso delle idee relative ai sistemi chiusi. Gli elementi di tale studio – come il gene, la cellula, l'organo, l'organismo, l'organismo indipendente, l'individuo, la società ecologica, il collettivo, ecc. – sono considerati inizialmente come sistemi potenzialmente isolati. Si esamineranno ora in modo molto generale alcune loro analogie e differenze con i sistemi fisici.

Le idee relative alla descrizione nello spazio delle fasi degli stati stabili, e spesso relative anche alle caratteristiche termodinamiche del sistema, sono applicate con successo quando il numero dei gradi di libertà osservati non è grande, oppure quando esso è grande ma i gradi di libertà sono abbastanza omogenei da rendere possibile un'osservazione statistica. Tali sono i modelli delle associazioni ecologiche secondo Volterra [1931] e i suoi seguaci, i modelli economici, le analisi statistiche della società secondo un parametro che interessi le compagnie di assicurazione e le attività politiche. La possibilità d'isolare un piccolo numero di gradi di libertà con la «compressione» dei rimanenti è largamente applicata agli esperimenti fisiologici e psicologici che permettono di analizzare con precisione la canalizzazione della reazione, cioè le traiettorie del sistema in uno spazio delle fasi di grandissima dimensione, costrette a trovarsi in un sottospazio di dimensione piccola in virtú di limitazioni esterne. Questo è il tipico metodo di lotta contro la principale instabilità dei sistemi biologici ricordata sopra, che permette di studiare separatamente i diversi canali concepibili del comportamento a spese dell'improvvisa limitazione di questa instabilità («cani di Pavlov»). Le leggi di conservazione o le loro varianti approssimative possono essere le conseguenze delle limitazioni fisiche corrispondenti: l'energetica, le risorse economiche, le potenzialità delle vie di comunicazione e dei mezzi di trasporto limitano in gran parte l'ambito del possibile per una società ecologica, o le nazioni in stato di guerra e pace.

Esiste tuttavia per lo meno una proprietà generale dei sistemi qui osservati e che non trova quasi corrispondenza nei comuni modelli della fisica. Questa proprietà limita l'attività del sistema biologico in rapporto al mondo esterno. L'apparato genetico della cellula è molto fortemente isolato da influssi esterni; la sua funzione consiste, anzi, nell'organizzazione attorno a sé della struttura complessa di un organismo vivente. L'organismo indipendente umano si difende con tutti i mezzi dagli influssi dell'ambiente esterno che escono dai limiti di un piccolo intervallo di temperatura, pressione, accelerazione, per la conservazione dell'omeostasi interna. Tuttavia esso compie una ricostruzione

molto precisa dell'ambiente del suo habitat. L'attività di un fermento che affretta in modo selezionato una reazione fra le molte possibili nelle sue vicinanze, e quella di una persona che ha costruito la cattedrale di Chartres o una macchina calcolatrice, hanno molto in comune tra di esse, e quasi niente in comune con il comportamento dei sistemi standard dei quali si occupa la fisica. La strutturazione direzionale della descrizione nello spazio delle fasi delle sue immediate vicinanze è, insieme al suo isolamento da influssi inversi, la proprietà piú caratteristica dei sistemi che interessano il biologo, lo psicologo o lo storico. Il problema della comprensione di tale comportamento, probabilmente, sarà sempre la massima sfida all'intelletto, tendente a capire la sua dipendenza dall'universo e la dipendenza dell'universo da sé. [JU. I. M.].

Feynmann, R. P., e Hibbs, A. R.
1965 *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York.

Rees, M.; Ruffini, R.; e Wheeler, J. A.
1975 *Black Holes, Gravitational Waves and Cosmology: An Introduction to Current Research*, Gordon and Breach, New York.

Thom, R.
1974 *Modèles mathématiques de la morphogénèse*, Union générale d'éditions, Paris.

Volterra, V.
1931 *Leçons sur la théorie mathématique de la lutte pour la vie*, Gauthier-Villars, Paris.

Weisskopf, V. F.
1970 *Modern Physics from an Elementary Point of View. Lectures Given in the Summer Vacation Program 1969*, Cern, Genève.

La nozione di dipendenza/indipendenza è un'idea metateorica generale con un amplissimo spettro di realizzazioni concrete. Come si è visto, si tratta di individuare un certo nucleo di concetti collegati alla descrizione della coppia dipendenza/indipendenza all'interno di alcuni modelli (cfr. **modello**) della **fisica** teorica. Dato che questa scienza utilizza la matematica come linguaggio di lavoro, i concetti generali di relazione e di funzione (cfr. **funzioni**) sono analizzati in rapporto alla teoria degli insiemi (cfr. **insieme**), la dipendenza lineare, quella algebrica e funzionale (cfr. **divisibilità**, **razionale/algebrico/trascendente**) e specificazioni piú complesse del concetto (cfr. **dualità**), fino alla dipendenza nei gruppi non commutativi e al «problema dell'identità delle parole» (cfr. **algoritmo**). Si pongono in luce i legami con fenomeni generali d'indecidibilità (cfr. **assioma/postulato, logica**) e questioni di dipendenza nel campo della **probabilità**.

L'astrazione che piú di ogni altra esprime nella **scienza** moderna l'idea di dipendenza/indipendenza è la concezione dei sistemi chiusi (cfr. **sistema**), concetto fondamentale per l'elaborazione di «leggi della natura» (cfr. **legge**), con le necessarie distinzioni fra il supporto discreto delle formule matematiche e la componente infinita, continua, variabile delle condizioni iniziali e della condizione al contorno (cfr. **analisi/sintesi**). Il principio di minima azione (cfr. **differenziale**, **variazione**) è di fondamentale importanza nella descrizione matematica dei sistemi isolati e viene illustrato con un esempio già trattato in **continuo/discreto** adatto a rappresentare i concetti principali dei sistemi chiusi. Questi concetti vengono poi riproposti nella loro generalità con l'aiuto di

strutture geometriche e topologiche (cfr. **geometria e topologia, applicazioni**). Si pongono in luce i legami tra le leggi di conservazione (cfr. **conservazione/invarianza**) fondamentali e la **simmetria** dei sistemi, e si esamina da questo punto di vista la teoria della **relatività.**

L'esempio dell'oscillatore lineare mostra come la descrizione matematica della causalità (cfr. **causa/effetto**) sia legata alla comparsa di funzioni analitiche. Un ulteriore esempio di sistemi chiusi è l'**universo**, descritto come modello cosmologico (cfr. **teoria/modello, caos/cosmo**) attraverso il concetto di varietà quadrimensionale con metrica pseudoriemanniana.

Il concetto di dipendenza è formulato anche in riferimento alla meccanica quantistica (cfr. **atomo e molecola, particella**) e alle interazioni termodinamiche (cfr. **entropia, equilibrio/squilibrio, controllo/retroazione**).

L'importanza della stabilità (cfr. **stabilità/instabilità**) è implicita nella considerazione dei sistemi isolati, mentre l'instabilità di alcuni regimi critici di comportamento è alla base degli effetti di amplificazione di tipo « meccanismi di difesa » (cfr. **catastrofi**).

Divisibilità

In aritmetica elementare 'divisibilità' di un numero naturale a per un numero naturale b significa l'esistenza di un numero naturale c tale che $a=bc$. Nella teoria del continuo dei numeri reali la divisibilità illimitata interviene quale possibilità, per ogni numero reale α e per ogni numero naturale b, di costruire un numero reale β tale che $b\beta=\alpha$. Nel linguaggio geometrico ciò significa che ogni segmento può essere diviso in un qualsiasi numero di parti uguali. In generale la divisione è l'operazione (parziale) inversa della moltiplicazione – classicamente, moltiplicazione per numeri interi –, operazione quest'ultima che equivale a sommare piú volte una stessa quantità e che è ammissibile per quantità del genere piú diverso (lunghezze, aree, volumi, pesi, ecc.). Dopo che la comprensione delle proprietà astratte della moltiplicazione ebbe permesso d'individuare il prodotto quale struttura che ammette diverse realizzazioni, si cominciò a chiamare divisione l'operazione inversa del prodotto negli anelli, nei monoidi, nei moduli, ecc. Là dove la divisione non è sempre possibile, la divisibilità diventa una relazione importante, non banale. Su un piano piú generale, alla teoria della divisibilità si può far risalire una parte notevole dell'algebra (commutativa), nella quale si studia in forma astratta l'interazione tra somma e prodotto. Infine, sin dai tempi della filosofia naturale dell'antica Grecia viene discusso il problema della divisibilità della materia, dello spazio, del tempo e la correlazione tra le proprietà del continuo geometrico divisibile e omogeneo e le proprietà della realtà fisica. A ciò si riconducono, in particolare, tutte le teorie atomistiche.

Questo articolo è dedicato prevalentemente agli aspetti algebrici e di teoria dei numeri della divisibilità. Nella sua parte elementare la teoria della divisibilità è contemporaneamente cosí semplice e cosí ricca di contenuto che si riporteranno qui alcuni risultati fondamentali con le dimostrazioni. Queste, nella maggior parte dei casi, si basano sul famoso algoritmo di Euclide per la divisione con resto, che è applicabile non solo a coppie di numeri interi, ma anche a qualsiasi coppia di numeri reali, o segmenti. In questa variante geometrica la divisione con resto non è altro che l'idealizzazione del procedimento fisico di misura di lunghezze con una unità di misura rigida. Perciò una successione di divisioni con resto è l'anello di congiunzione tra i procedimenti per *contare* grandezze discrete e *misurare* grandezze continue. Qui è d'importanza fondamentale l'assioma di Archimede, secondo il quale una quantità, per quanto piccola, ripetuta un numero sufficiente di volte, fornisce una quantità grande a piacere. (Nella moderna teoria della divisibilità si è condotti a considerare altre nozioni di «piccolezza», per le quali l'assioma di Archimede non è piú vero; cfr. § 5).

In tal modo nella teoria dei numeri reali si introduce lo strumento delle frazioni continue e lo studio della struttura fine del continuo (cfr. § 3).

Nell'aritmetica dei numeri interi otteniamo la teoria del massimo comun

divisore e della scomposizione in fattori primi. I risultati (in misura minore i metodi) sono un modello per la teoria dei numeri algebrici e, con la chiarificazione della nozione di ideale, per i concetti algebrico-geometrici. In questi ultimi si sfrutta il profondo parallelismo tra numeri e funzioni, divisibilità per un numero primo e annullamento in un punto; parallelismo che nel presente articolo sarà solo toccato nell'esempio elementare numeri interi - polinomi.

Infine sono da notare i problemi della geometria diofantea (soluzione di equazioni indeterminate in numeri interi), la quale, da secoli, è sorgente di problemi ed è la pietra di paragone dei metodi della teoria dei numeri. Nei suoi risultati entrambi gli aspetti della divisibilità, numerico e funzionale, sono di pari importanza.

1. *Teoria elementare della divisibilità e numeri primi.*

1.1. Divisione con resto.

Siano $a, b \neq 0$ due numeri reali. Esistono allora un numero intero q e un numero reale r tali che $a=bq+r$, $0 \leq r < |b|$. Di piú, q e r sono unici. Il numero q si dice quoziente (incompleto, o parziale) della divisione di a per b; r si dice resto; b è il divisore, a il dividendo.

Non è difficile dedurre geometricamente l'esistenza e l'unicità di q, r. Per a, b positivi, q è il massimo numero di volte che il segmento di lunghezza b è interamente contenuto nel segmento di lunghezza a, mentre r è la parte del segmento a, minore di b, che rimane dopo di ciò. (Per $a<0$, $b>0$ il numero $q<0$ è il massimo con la proprietà $a \geq bq$, cioè $|a| \leq b|q|$: il segmento di lunghezza $|a|$ è ricoperto da $|q|$ segmenti di lunghezza b, ma non da un numero inferiore).

Per la teoria della misura è essenziale che a, b, r possano designare quantità di natura indeterminata: la possibilità di confrontarle per mezzo della divisione con resto, o, secondo Eudosso, mediante l'approssimazione di rapporti di numeri interi, segue da alcuni assiomi che qui non formuleremo esplicitamente.

Se $r=0$, a è interamente divisibile per b; il risultato della divisione, q, è la misura della quantità a, quando come unità di misura si assuma la quantità b. Le quantità a, b (per esempio, segmenti) sono commensurabili se, per un opportuno intero p, la quantità pa è interamente divisibile per b. Affermazione equivalente: il rapporto tra le misure di a e di b è razionale. Quando a e b sono interi ed a è interamente divisibile per b, b si dice divisore di a; notazione $b \mid a$.

1.2. Algoritmo di Euclide.

Sia, come sopra, $a=bq+r$, $0 \leq r < |b|$. Se $r>0$, si esegua la divisione con resto di b per r: $b=rq_1+r_1$. Se $r_1>0$, si esegua la divisione con resto di r per r_1: $r=r_1q_2+r_2$; e cosí via. Questo procedimento di divisioni successive con resto si chiama algoritmo di Euclide: il divisore e il resto del passo precedente diventano rispettivamente dividendo e divisore al passo successivo. L'algoritmo

si ferma se e quando il resto è uguale a zero. Per questo è necessario e sufficiente che a e b siano commensurabili. Tutta la successione dei quozienti parziali della divisione di a per b definisce lo sviluppo di ab^{-1} in una frazione continua (cfr. § 3.2).

In particolare, l'algoritmo si ferma sempre se a e b sono numeri interi. L'ultimo resto non nullo nella divisione di a per b è il massimo comun divisore di a e b: il massimo numero reale per il quale siano interamente divisibili a e b. Questa nozione è della piú grande importanza quando a e b sono interi. Prima di studiarla nei particolari, diciamo ancora alcune parole sull'aspetto computazionale dell'algoritmo della divisione.

Questo aspetto si esaurisce nella descrizione del procedimento fisico idealizzato che, nella variante geometrica, consiste nel riportare un segmento su di un altro segmento. Nella variante algebrica l'algoritmo della divisione diventa realmente possibile solo in un sistema di numerazione sufficientemente sviluppato. Già la scrittura di un numero intero, ad esempio nel sistema decimale, contiene l'informazione sulla divisione con resto di questo numero per una potenza di dieci. L'algoritmo della divisione nella scrittura decimale (quello che si impara nei primi anni di scuola) è già di per sé un'importante scoperta di teoria dei numeri, ma secondaria rispetto al significato della divisione, che è indipendente dal sistema di numerazione. Da questo punto di vista il linguaggio geometrico dei Greci è piú adeguato all'aritmetica teorica che non i comodi sistemi di calcolo dei tempi piú recenti.

1.3. Massimo comun divisore.

Siano $a, b \neq 0$ due numeri interi. Un numero intero $d > 0$ si dice divisore comune di a, b se $d \mid a$ e $d \mid b$. Un numero intero $d > 0$ si dice massimo comun divisore di a e b se esso è divisibile per ogni loro divisore comune. Se un tale numero esiste, esso è unico.

Mostriamo che il massimo comun divisore di a e b esiste e si rappresenta nella forma $ax+by$ con opportuni numeri interi x, y. Infatti, indichiamo con d il piú piccolo numero intero positivo della forma $ax+by$, ad esempio $d = ax_0 + by_0$, e mostriamo che a e b sono divisibili per d. Dividiamo a per ax_0+by_0 con resto; otteniamo $a = (ax_0+by_0)q+r$. Da cui si ricava $r = a(1-qx_0) + b(-qy_0)$ con $0 \leq r \leq d-1$. Cioè r si rappresenta nella forma $ax+by$, non è negativo ed è minore di d; perciò $r=0$, cosí che a è divisibile per d. Analogamente si dimostra che b è divisibile per d.

D'altro lato, se $d_0 \mid a$ e $d_0 \mid b$, d_0 divide qualsiasi numero della forma $ax+by$; in particolare d. Perciò d è il massimo comun divisore di a e b. Non è difficile verificare che la definizione data in questo punto coincide con quella data nella descrizione dell'algoritmo di Euclide, cioè che l'ultimo resto non nullo effettivamente coincide con d. A partire da d e percorrendo a ritroso i passi di questo algoritmo, si trova la soluzione esplicita dell'equazione $d = ax+by$. Notazione abbreviata: $d = \text{MCD}(a,b)$, oppure $d = (a,b)$.

Se $(a,b) = 1$, i numeri a e b si dicono primi tra loro.

1.4. Numeri primi.

Un numero intero $p > 1$ si dice primo se ogni suo divisore positivo è 1, oppure p. I primi numeri primi sono: 2, 3, 5, 7, 11, 13, ... Ogni numero intero a è divisibile almeno per un numero primo: il suo piú piccolo divisore, maggiore di 1, è primo (altrimenti un divisore di questo divisore, diverso da esso stesso e da 1, sarebbe piú piccolo).

Da qui segue in primo luogo che ogni numero intero $a > 1$ si scompone in un prodotto finito di numeri primi: ciò è vero per $a = 2$ e, se ciò è già stato dimostrato per tutti i numeri minori di a, ne segue che a si scompone in $a = p \times$ $\times$(decomposizione di ap^{-1}), dove p è un divisore primo di a.

In secondo luogo segue che i numeri primi sono infiniti. Infatti, per ogni insieme finito di numeri primi $p_1, \ldots, p_r$, ogni divisore primo del numero $N = p_1 \ldots p_n + 1$ è diverso da $p_1, \ldots, p_n$, poiché il resto della divisione di N per $p_1, \ldots, p_n$ è uguale a 1 e non a 0.

1.5. Unicità della decomposizione in fattori primi.

Siano $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, ... tutti i numeri primi scritti in ordine crescente. Sia $a \geq 1$ un numero intero e $a = p_1^{v_1} p_2^{v_2} \ldots p_k^{v_k}$ la decomposizione di a in un prodotto (di potenze) di numeri primi (per $a = 1$ ogni v_i è nullo). Mostriamo che questa decomposizione è univocamente definita per induzione su a.

Per $a = 1$ è evidente. Sia $a \geq 2$. Consideriamo ancora un'altra decomposizione: $a = p_1^{w_1} p_2^{w_2} \ldots p_k^{w_k}$. Supponiamo che queste decomposizioni siano diverse e giungeremo a una contraddizione.

Scegliamo i tale che $v_i \neq w_i$. Se $v_i > 0$ e $w_i > 0$, dividendo a per p_i^v, dove v è il piú piccolo dei numeri v_i e w_i, otteniamo due diverse decomposizioni del numero ap_i^{-v}, in contraddizione con l'ipotesi che per i numeri minori di a la decomposizione sia unica.

Se invece, diciamo, $v_i = 0$, mentre $w_i > 0$, si ricava che $a = p_1^{v_1} \ldots p_k^{v_k}$ è divisibile per p_i pur essendo $v_i = 0$. Dimostriamo che ciò è impossibile per induzione su $v_1 + \ldots + v_k > 0$. Se $v_1 + \ldots + v_k = 1$, esattamente un solo esponente v_j è uguale a 1; gli altri sono 0. Allora $p_1^{v_1} \ldots p_k^{v_k} = p_j$, $j \neq i$, ma p_j non è divisibile per $p_i \neq p_j$, poiché p_j è un numero primo e $p_i \neq 1$.

Infine supponiamo che $v_1 + \ldots + v_k > 1$ e che per valori minori della somma degli esponenti sia già stato dimostrato che $p_1^{w_1} \ldots p_k^{w_k}$ non è divisibile per p_i se $w_i = 0$. Allora $a = p_1^{v_1} \ldots p_k^{v_k}$ si può mettere nella forma $a = bc$, dove $b > 0$, $c > 0$ e b, c non sono divisibili per p_i. Dimostriamo che allora neanche a è divisibile per p_i. Infatti, se b, c non sono divisibili per p_i, essi sono primi con p_i. Siano x, y, x', y' numeri interi tali che $bx + p_i y = 1$ e $cx' + p_i y' = 1$. Moltiplicando queste uguaglianze troviamo: $bcxx' + p_i cx'y + p_i bxy' + p_i^2 yy' = 1$. Se si suppone che p_i divida bc, si ricava che p_i divide 1 e ciò è una contraddizione. Ciò conclude la dimostrazione dell'unicità. Quale corollario non è difficile ottenere che, se $a = \prod p_i^{v_i}$ e $b = \prod p_i^{w_i}$, si ha: $\text{MCD}(a,b) = \prod p_i^{\min(v_i, w_i)}$.

1.6. Numeri razionali.

Dal teorema sull'esistenza e l'unicità della decomposizione dei numeri naturali in fattori primi segue che ogni numero razionale (cioè ogni rapporto tra interi) si rappresenta in modo univoco sotto forma di frazione irriducibile a/b, oppure $-(a/b)$, dove $a \geq 0$, $b > 0$ sono naturali con $(a,b)=1$; oppure anche nella forma $\pm\prod p_i^{v_i}$, dove, come sopra, $p_1=2$, $p_2=3$, $p_3=5$, ... sono primi, mentre v_i sono esponenti interi qualsiasi, dei quali solo un numero finito sono diversi da zero.

1.7. Distribuzione dei numeri primi nella serie naturale.

Un metodo semplice per formare la tabella dei numeri primi che non superano N è il seguente. Da una tabella di tutti i numeri interi da 1 a N si cancella 1 (passo zero); poi si cancella ogni numero pari ad eccezione di 2 (primo passo); poi ogni multiplo di 3 ad eccezione di 3 (secondo passo) e cosí via. Se al k-esimo passo è stato cancellato ogni multiplo di p_k, ad eccezione di p_k, al passo $(k+1)$-esimo bisogna prendere quale p_{k+1} il piú piccolo numero, dopo p_k, non ancora cancellato e cancellare tutti i multipli di p_{k+1}, ad eccezione di p_{k+1}. I numeri non cancellati $p_1, p_2, \ldots$ sono appunto tutti i numeri primi. Questo metodo di setacciare tutti i numeri composti si chiama crivello di Eratostene. Le tabelle dei numeri primi rivelano una straordinaria irregolarità nella loro distribuzione nella successione naturale. Tutte le nuove generazioni di matematici compiono tentativi per trovare delle regolarità in questo apparente caos. Accenniamo ad alcuni risultati rimarchevoli, benché la loro dimostrazione esca dall'ambito della teoria elementare della divisibilità e di questo articolo.

Il numero $\pi(N)$ dei numeri primi che non superano N asintoticamente si comporta come $N/\log N$; l'errore non supera $\operatorname{cost} N/\log^2 N$. Un'approssimazione assai migliore è data dal cosiddetto logaritmo integrale $\operatorname{li} N = \int_2^N (1/\log x)\,dx$. Il vero ordine dell'errore non è noto: si congettura che esso non sia superiore a $cx^{1/2+\varepsilon}$, con $\varepsilon > 0$ qualsiasi e $c = c(\varepsilon)$ una costante opportuna. Si ha una formula esatta per $\pi(N) - \operatorname{li} N$, la quale si esprime come somma infinita rispetto agli zeri della funzione zeta di Riemann $\zeta(s)$: funzione meromorfa nella variabile complessa s, che viene definita per $\operatorname{Re}(s) > 1$ mediante la serie $\sum_{n=1}^{\infty} \frac{1}{n^s}$ e che poi viene estesa a tutto il piano complesso con uno dei noti e numerosi metodi. I suoi zeri non banali giacciono nella striscia critica $0 < \operatorname{Re}(s) < 1$. La congettura di Riemann, finora non dimostrata, afferma che essi giacciono sulla retta $\operatorname{Re}(s) = 1/2$. Questa congettura è equivalente alla stima $|\pi(N) - \operatorname{li} N| \leq \operatorname{cost} \cdot N^{1/2+\varepsilon}$ per ogni $\varepsilon > 0$. Il legame tra i numeri primi e gli zeri di $\zeta(s)$ è una delle illustrazioni piú famose e sorprendenti della sintesi tra continuo e discreto nella teoria dei numeri.

Se $(a,b)=1$, in ogni progressione aritmetica $ax+b$ vi sono infiniti numeri primi (Dirichlet). Inoltre, asintoticamente il numero di tali numeri $\leq N$ è uguale a $(1/\varphi(a))\operatorname{li} N$, dove $\varphi(a)$ è il numero degli interi c tali che $1 \leq c \leq a-1$ e $(c,a)=1$ (cfr. § 2). Pertanto, asintoticamente, i numeri primi sono distribuiti uniformemente nelle progressioni. Tuttavia le differenze fra quantità di numeri primi in due progressioni con uno stesso a rivelano irregolarità di comportamento sottili e ancora poco comprese. Cosí, nei limiti di tabelle molto lunghe di numeri primi, quelli della forma $4n+3$ sono sempre piú di quelli della forma $4n+1$. Ciò nondimeno è dimostrato che la differenza di queste due quantità cambia segno infinite volte.

Si comportano irregolarmente anche le differenze di numeri primi successivi. Dall'asintotica della loro quantità è chiaro che $p_{k+1} - p_k$ deve oscillare intorno a $\log k$, ma l'ampiezza di queste oscillazioni è molto grande e non sono noti né il suo estremo superiore, né il suo estremo inferiore. Si congettura che infinite volte si ha $p_{k+1} - p_k = 2$, ma questo non è dimostrato (problema dei gemelli).

Fin dai tempi antichi si tentò di trovare una «formula esplicita», cioè una funzione di forma semplice, che, per valori interi dell'argomento, desse solo numeri primi; tutti, o almeno una parte. Tali tentativi condussero rapidamente a una matematica molto ricca. Una soluzione inattesa di tale questione si è ottenuta non molto tempo fa. Si è dimostrata l'esistenza di un polinomio in più variabili $P(x_1, \ldots, x_n)$ a coefficienti interi tale che i suoi valori positivi nei punti interi $x_1, \ldots, x_n \in \mathbf{Z}$ sono tutti e soli i numeri primi, eventualmente con ripetizioni e in un ordine sconosciuto. La stessa cosa è vera, tra l'altro se, al posto della successione dei primi, si considera una qualsiasi successione computabile di numeri naturali, cioè l'insieme dei valori positivi di ogni funzione ricorsiva (algoritmicamente calcolabile) da $\mathbf{Z}_+$ a $\mathbf{Z}_+$. Per via della generalità di questo magnifico risultato, finora non è stato ancora possibile usarlo per ottenere nuove concrete notizie sui numeri primi.

1.8. Numeri perfetti e altro.

Alla nozione di divisibilità è legata una classe di funzioni aritmetiche, come $\tau(n) = \sum_{d|n} 1$ (numero di divisori di n), $d(n) = \sum_{d|n} d$ (somma dei divisori di n), ecc. È nota tutta una serie di problemi risolti e non risolti sul loro comportamento.

Un numero n si dice perfetto se $d(n) = 2n$. Negli *Elementi* di Euclide è data la costruzione di tutti i numeri perfetti pari. La questione dell'esistenza di numeri perfetti dispari finora è ancora aperta; tale questione, a differenza di alcuni problemi di teoria dei numeri elementari e molto difficili, non ha fatto sorgere nuove nozioni o idee che abbiano un qualche interesse.

Nella teoria analitica dei numeri si studia il comportamento in media di funzioni quali $\tau(n)$, o $d(n)$. Per esempio, asintoticamente abbiamo $\sum_{n \leq N} \tau(n) \sim N \log N$. Questa stima e stime piú precise si possono ottenere osservando che $\sum_{n \leq N} \tau(n)$ è il numero delle soluzioni in numeri naturali della disequazione $xy \leq N$, cioè il numero dei punti coordinate intere nel primo quadrante sotto l'iperbole

$xy = N$, approssimativamente uguale all'area di questa porzione di piano (limitata evidentemente dalle disuguaglianze $x \leq N$ e $y \leq N$). In tale genere di problemi di solito è relativamente facile ottenere il termine principale dello sviluppo asintotico, mentre problemi sottili sono legati alle stime e alle formule esatte dei resti.

2. *Congruenze.*

Sia $m > 0$ un numero intero. Allora i resti della divisione dei numeri interi per m formano un sistema chiuso rispetto alla somma e al prodotto e al passaggio successivo ai resti. Questo sistema si chiama «anello delle classi dei resti modulo m». Il suo studio ha condotto a molte importanti scoperte di teoria dei numeri e ha favorito la formazione di una serie di fondamentali nozioni dell'algebra.

2.1. Definizione dell'anello delle classi dei resti $\mathbf{Z}/m\mathbf{Z}$.

L'anello delle classi dei resti $\mathbf{Z}/m\mathbf{Z}$ è un insieme con due operazioni, $+$ e $\times$, e con due elementi particolari, $0 \bmod m$ e $1 \bmod m$.

I suoi elementi sono tutte le progressioni aritmetiche $a + m\mathbf{Z} = \{a + mx \mid x \in \mathbf{Z}\}$, le quali sono sottoinsiemi di $\mathbf{Z}$. Al posto di $a + m\mathbf{Z}$ scriveremo $a \bmod m$, oppure $\bar{a}$, quando m è fissato e si può tralasciare di ricordarlo. Il numero m è il modulo, $\bar{a}$ è la classe dei resti del numero a, a è un rappresentante della classe $\bar{a}$. I numeri $\{0, 1, 2, \ldots, m-1\}$ formano un sistema completo di rappresentanti, uno per ogni classe se $m \geq 1$. Per $m = 0$ abbiamo $\mathbf{Z}/0\mathbf{Z} = \mathbf{Z}$.

Si dice somma $\bar{a} + \bar{b}$ la classe $\overline{a+b}$; si dice prodotto $\bar{a} \times \bar{b}$, oppure semplicemente $\bar{a}\bar{b}$, la classe $\overline{ab}$. Le classi $\bar{a} + \bar{b}$ e $\bar{a}\bar{b}$ non dipendono dalla scelta dei rappresentanti $a \in \bar{a}$, $b \in \bar{b}$, ciò che si verifica con un semplice calcolo. Al posto di $\bar{a}_1 = \bar{a}_2$ si scrive $a_1 \equiv a_2 \bmod m$.

Rispetto alla somma $+$ l'insieme $\mathbf{Z}/m\mathbf{Z}$ forma un gruppo abeliano (commutativo) con elemento neutro $\bar{0}$. Rispetto al prodotto, $\mathbf{Z}/m\mathbf{Z}$ forma un monoide commutativo con unità $\bar{1}$: $\bar{1}\,\bar{a} = \bar{a}$ per ogni $\bar{a}$ e $\bar{a}\bar{b} = \bar{b}\bar{a}$. Il prodotto è distributivo rispetto alla somma. Perciò $\mathbf{Z}/m\mathbf{Z}$ è un anello (per le definizioni di gruppo, monoide, anello si veda l'articolo «Applicazioni» della presente *Enciclopedia*, § 3). Esso consiste di m elementi per $m \neq 0$ e coincide con $\mathbf{Z}$ per $m = 0$.

L'applicazione $\mathbf{Z} \to \mathbf{Z}/m\mathbf{Z} : a \mapsto \bar{a}$ è un omomorfismo di anelli.

Il gruppo additivo dell'anello $\mathbf{Z}/m\mathbf{Z}$ consiste degli elementi $\bar{0}$, $\bar{1}$, $\bar{1} + \bar{1}$, ..., $\bar{1} + \bar{1} + \ldots + \bar{1}$ ($m - 1$ volte), cioè è un gruppo ciclico di ordine m (cfr. § 4).

Il monoide moltiplicativo di $\mathbf{Z}/m\mathbf{Z}$ non è un gruppo per $m \neq 1$: la classe $\bar{0}$ non è invertibile.

Affinché l'anello $\mathbf{Z}/m\mathbf{Z}$ sia un corpo è necessario e sufficiente che ogni classe $\bar{a} \neq \bar{0}$ sia invertibile, cioè che l'insieme $\mathbf{Z}/m\mathbf{Z} - \{\bar{0}\}$ sia un gruppo rispetto al prodotto. Mostriamo che questa condizione insieme con $\bar{0} \neq \bar{1}$ equivale a richiedere che m sia un numero primo.

Se m non è primo, sia $m = ab$, dove $1 < a, b < m$. Allora $\bar{a} \neq \bar{0}$, $\bar{b} \neq \bar{0}$, ma $\bar{a}\bar{b} = \bar{0}$. Perciò $\bar{a}$ e $\bar{b}$ non sono invertibili e $\mathbf{Z}/m\mathbf{Z}$ non è un corpo.

Al contrario, sia m primo e $\bar{a} \neq \bar{0}$. Allora m non è divisibile per a; d'accordo con il § 1, a è primo con m ed esistono $x, y \in \mathbf{Z}$ tali che $ax + my = 1$. Da cui segue che $\bar{a}\bar{x} = \bar{1}$, cioè $\bar{x}$ è l'elemento inverso di $\bar{a}$.

2.2. La funzione di Eulero.

Le considerazioni dell'ultimo punto possono essere generalizzate al caso di un qualsiasi modulo m e ottenere il risultato seguente: se $(a, m) = 1$, la classe $\bar{a} = a \bmod m$ è invertibile. Al contrario, se $(a, m) = d > 1$, $\bar{a}$ non è invertibile poiché $\bar{a} \cdot \overline{md^{-1}} = \bar{0}$ e $\overline{md^{-1}} \neq \bar{0}$.

Indichiamo con $(\mathbf{Z}/m\mathbf{Z})^*$ l'insieme delle classi $\bar{a}$, $(a, m) = 1$. Esse formano un gruppo commutativo rispetto al prodotto. Il numero dei suoi elementi si indica con $\varphi(m)$ e questa si chiama funzione di Eulero (quale funzione di m). Se p è primo, $\varphi(p) = p - 1$; un calcolo diretto mostra che $\varphi(p^v) = p^{v-1}(p-1)$ per ogni $v \geq 1$. Considerazioni un po' piú complesse dànno la formula generale $\varphi(p_1^{v_1} \ldots p_k^{v_k}) = \prod_{i=1}^{k} p_i^{v_i - 1}(p_i - 1)$, dove p_i sono numeri primi diversi tra loro. In particolare, se $(m, n) = 1$, si ha $\varphi(mn) = \varphi(m)\varphi(n)$: la funzione di Eulero è moltiplicativa.

2.3. I teoremi di Fermat e di Eulero.

Il teorema di Eulero afferma che, se $(a, m) = 1$, $a^{\varphi(m)} \equiv 1 \bmod m$. Nel caso particolare di $m = p$ primo, si ha il (piccolo) teorema di Fermat: se a non è divisibile per p, $a^{p-1} \equiv 1 \bmod p$, da cui segue che per ogni $a \in \mathbf{Z}$, $a^p \equiv a \bmod p$.

Per dimostrare il teorema di Eulero consideriamo l'insieme $\{\bar{a}_1, \ldots, \bar{a}_{\varphi(m)}\}$ di tutte le classi invertibili. Mostriamo che, se $\bar{a}$ è invertibile, $\{\bar{a}\bar{a}_1, \ldots, \bar{a}\bar{a}_{\varphi(m)}\}$ coincide con $\{\bar{a}_1, \ldots, \bar{a}_{\varphi(m)}\}$ (quale insieme, cioè a meno di una permutazione). Infatti $\bar{a}\bar{a}_i$ è invertibile insieme con $\bar{a}$ e $\bar{a}_i$, e $\bar{a}\bar{a}_i \neq \bar{a}\bar{a}_j$ per $\bar{a}_i \neq \bar{a}_j$ (si moltiplichi per $\bar{a}^{-1}$ la supposta uguaglianza $\bar{a}\bar{a}_i = \bar{a}\bar{a}_j$). Perciò $\bar{a}\bar{a}_1, \ldots, \bar{a}\bar{a}_{\varphi(m)}$ consiste di $\varphi(m)$ classi invertibili diverse e quindi è $(\mathbf{Z}/m\mathbf{Z})^*$.

Da qui segue che $\prod_{i=1}^{\varphi(m)} \bar{a}_i = \prod_{i=1}^{\varphi(m)} \bar{a}\bar{a}_i = \bar{a}^{\varphi(m)} \prod_{i=1}^{\varphi(m)} \bar{a}_i$. Dividendo per la classe invertibile $\prod_{i=1}^{\varphi(m)} \bar{a}_i$, si ottiene: $\bar{a}^{\varphi(m)} = \bar{1}$.

2.4. Il teorema di Wilson.

Esso afferma che, se p è un numero primo, $(p-1)! + 1 \equiv 0 \bmod p$.

Per $p = 2$ è evidente. Sia $p \geq 3$. Se $\bar{a} = \bar{a}^{-1}$ in $\mathbf{Z}/p\mathbf{Z}$, $\bar{a} = \bar{1}$, oppure $\bar{a} = \overline{p-1}$. L'equazione $\bar{a}^2 - \bar{1} = \bar{0}$ non ha piú di due radici nel corpo $\mathbf{Z}/p\mathbf{Z}$, come del resto in ogni corpo. Ad eccezione di $\bar{1}$ e $\overline{p-1}$ le altre classi non nulle di residui mod p si dividono in $(p-3)/2$ coppie $(\bar{a}, \bar{a}^{-1})$ tali che il loro prodotto è $\bar{1}$. Perciò $\overline{(p-1)!} = \bar{1} \cdot \overline{(p-1)} = -\bar{1}$.

In realtà il teorema di Wilson ci dà una condizione necessaria e sufficiente affinché p sia primo: se m non è primo e d, $1<d<m$, divide m, $(m-1)!$ è divisibile per d e quindi $(m-1)!+1$ non è divisibile per d e, a maggior ragione, non lo è per m.

3. *Frazioni continue e approssimazioni razionali.*

3.1. Approssimazioni razionali.

La maggior parte dei numeri reali sono formazioni infinite: un numero può essere dato, per esempio, mediante le cifre decimali, che formano una successione infinita. Praticamente e teoricamente hanno molta importanza le approssimazioni razionali di un numero α, cioè quei numeri razionali (a/b) tali che la distanza $|\alpha-(a/b)|$ sia piccola. Stimiamo l'ordine di piccolezza che si può raggiungere in funzione del denominatore b.

È sufficiente considerare il caso $0\leq\alpha<1$. I numeri razionali con denominatore b dividono $[0, 1]$ in b intervalli $[i/b, (i+1)/b]$, $i=0, \ldots, b-1$, di lunghezza $1/b$. Il punto α capita in uno di essi e, indicando con a/b l'estremo di questo intervallo piú vicino a α, si ricava $|\alpha-(a/b)|<1/2b$. Tuttavia *alcuni* denominatori b, che dipendono da α, permettono di ottenere un'approssimazione notevolmente migliore: $|\alpha-(a/b)|<1/b^2$.

Per la dimostrazione consideriamo un numero intero $B\geq 1$ e tutte le possibili coppie (a, b) di numeri interi, $a\geq 0$, $b>0$, con $0\leq b\alpha-a<1$, $b\leq B$. Per ognuno dei $B+1$ valori di b vi è un a che verifica la condizione: il piú grande a per cui $a\leq b\alpha$. Perciò esiste i, $0\leq i\leq B-1$, tale che i due valori $b_1\alpha-a_1$ e $b_2\alpha-a_2$ capitano nell'intervallo $[i/B, (i+1)/B]$, poiché in tutto di tali intervalli se ne hanno B. Ponendo $b=b_1-b_2$ (se $b_1>b_2$) e $a=a_1-a_2$, otteniamo: $|b\alpha-a|<1/B$, da cui $|\alpha-(a/b)|<1/Bb\leq 1/b^2$.

Questo ragionamento è dovuto a Dirichlet.

Queste approssimazioni con i numeri razionali particolarmente buone sono state usate praticamente. Ecco alcuni esempi.

Il classico valore $22/7$ è una buona approssimazione di π:

$$|\pi-(22/7)|<1/7\cdot 100.$$

La costruzione dei modelli del sistema solare, nei quali il moto dei pianeti lungo le orbite viene realizzato mediante trasmissioni ad ingranaggi, richiede che si conoscano buone approssimazioni razionali dei rapporti tra i periodi di rivoluzione dei pianeti. I numeratori e denominatori di tali approssimazioni corrispondono ai numeri di denti degli ingranaggi, i quali, per questioni tecniche, non possono essere troppo grandi. Tali approssimazioni sono state trovate da Christian Huygens che costruí questi modelli.

L'esistenza nel sistema solare di grossi pianeti con il rapporto tra i loro periodi vicino a $2/5$ (Saturno-Giove) ha serie conseguenze cosmografiche: periodicamente (ogni ottocento anni) si ha una grande perturbazione del moto

medio dei pianeti, la quale può generare instabilità di varie dimensioni. È, questo, un fenomeno tipico dei modelli matematici della meccanica celeste («piccoli denominatori» nelle serie della teoria delle perturbazioni).

Infine, curiose questioni sulle buone approssimazioni sono legate alla storia della formazione della scala musicale di dodici suoni, o scala cromatica (Andreas Werckmeister, intorno al 1700).

Le frequenze fondamentali delle corde di un pianoforte accordato formano una successione finita di numeri soggetta ad alcune condizioni. Prima di tutto queste frequenze devono formare una progressione geometrica affinché la scala sia uniforme e ogni melodia possa iniziare con un qualsiasi tono (l'udito percepisce due successioni di suoni come se fossero uguali, a meno dell'altezza, se i rapporti tra le frequenze dei suoni corrispondenti sono costanti). Inoltre, insieme con ogni frequenza, la scala deve contenere quella doppia e quella metà (nei limiti dell'intervallo delle frequenze di un dato strumento). Ciò è legato al fatto che la frequenza doppia forma l'armonica predominante nella corda, complementare alla nota fondamentale, e, per le combinazioni armoniche, essa deve far parte della scala.

L'intervallo da ogni frequenza f a $2f$ forma un'ottava. Esso è troppo grande e dev'essere suddiviso in un certo numero k di parti piú piccole (che si chiamano semitoni). Il rapporto tra due semitoni successivi in una scala uniforme, perciò, deve essere uguale a $2^{1/k}$. D'altro lato, le condizioni acustiche per la possibilità di combinazioni armoniche richiedono che i rapporti tra le frequenze nella scala siano numeri razionali con numeratori e denominatori piccoli: tali sono i rapporti fra le frequenze delle armoniche predominanti in un dato tono. Poiché per ogni k il numero $2^{1/k}$ è irrazionale, queste esigenze sono incompatibili. Bisogna perciò giungere a un compromesso e rinunziare agli intervalli puri. La scelta di $k=12$ è legata all'esistenza di una buona approssimazione razionale, $7/12$, di $\log_2 3/2$, cioè della buona approssimazione, $3/2$ (rapporto fra le sequenze in una quinta), di $2^{7/12}$: $7/12\sim 0{,}583$, $\log_2 3/2\sim 0{,}585$, cosí che nella prima ottava (262-524 hertz) la differenza tra le frequenze di un sol ideale e reale è inferiore a 1 hertz.

E cosí si sceglie $k=12$; la frequenza di $(3/2)f_0$ viene sostituita con $2^{7/12}f_0$; per gli altri toni della scala uniforme, $2^{i/12}$, pure si scelgono buone approssimazioni razionali: $(9/8)f_0\sim 2^{1/6}f_0$ (re); $(5/4)f_0\sim 2^{1/3}f_0$ (mi); $(4/3)f_0\sim 2^{5/12}f_0$ (fa); $(3/2)f_0\sim 2^{7/12}f_0$ (sol); $(5/3)f_0\sim 2^{1/4}f_0$ (la); $(15/8)f_0\sim 2^{11/12}f_0$ (si).

Prima di Werckmeister gli strumenti venivano accordati secondo intervalli puri (quinte, terze). Molti, e tra essi Diderot, furono avversi alla nuova scala, nella quale non vi sono intervalli puri. Solo dopo il *Clavicembalo ben temperato* di Bach essa ottenne un generale riconoscimento. I principî acustici, di teoria dei numeri, che in essa sono fusi, hanno definito in larga misura il carattere della musica europea dei tempi moderni.

Un metodo sistematico per trovare buone approssimazioni razionali è fornito dalla tecnica delle frazioni continue.

3.2. Frazioni continue.

Sia $a \geq 0$ un numero reale. Applichiamo l'algoritmo di Euclide, descritto nel § 1, alla coppia $(\alpha, 1)$ e indichiamo i quozienti parziali successivi con $a_1, a_2, a_3, \ldots$ (questi sono numeri naturali). Pertanto $\alpha = a_1 + r_1$, $0 \leq r_1 < 1$; $1 = a_2 r_1 + r_2$, $0 \leq r_2 < r_1$ (se $r_1 \neq 0$); $r_1 = a_3 r_2 + r_3$, $0 \leq r_3 < r_2$ (se $r_2 \neq 0$) e cosí via. Perciò successivamente si trova:

$$\alpha = a_1 + r_1 = a_1 + \cfrac{1}{a_2 + \cfrac{r_2}{r_1}} = a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{r_3}{r_2}}} = \ldots = a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\ldots \cfrac{1}{a_k}}}}$$

L'ultima uguaglianza si verifica se $r_k = 0$, cioè nel caso di α razionale. Scriviamo l'ultima espressione nella forma $\alpha = [a_1; a_2, \ldots, a_k]$. Essa si chiama frazione continua di lunghezza k con i quozienti parziali $a_1, \ldots, a_k$.

Se α è irrazionale, la successione dei quozienti parziali (a_n) non finisce. I numeri corrispondenti $\alpha_n = [a_1; a_2, \ldots, a_n]$ si dicono frazioni ridotte di α. Il limite di α_n esiste ed è uguale ad α. Perciò si può scrivere $\alpha = [a_1; a_2, a_3, \ldots]$, sottintendendo una frazione infinita. Quale (a_i) si può avere una qualsiasi successione di numeri naturali.

Definiamo i polinomi universali P_n, Q_n nelle variabili $a_1, a_2, \ldots$, a coefficienti interi, primi tra loro, mediante le formule $P_k/Q_k = [a_1; a_2, \ldots, a_k]$ per $k \geq 1$; $P_0 = 1$, $Q_0 = 0$. Per induzione su k non è difficile dimostrare una serie d'identità per P_n, Q_n. Per esempio, $P_n Q_{n-1} - P_{n-1} Q_n = (-1)^n$, $n \geq 1$, da cui

$$\frac{P_n}{Q_n} - \frac{P_{n-1}}{Q_{n-1}} = \frac{(-1)^n}{Q_{n-1} Q_n}.$$

D'altro lato, è facile verificare che le ridotte $\alpha_n = p_n/q_n$ di ogni $\alpha \geq 0$ con indice n pari sono maggiori di α, mentre quelle con indice dispari sono minori (ad eccezione dell'ultima, che è uguale ad α, se α è razionale). Perciò

$$\left| \alpha - \frac{p_{n-1}}{q_{n-1}} \right| \leq \frac{1}{q_{n-1} q_n} < \frac{1}{q_{n-1}^2},$$

cosí che le frazioni ridotte sono buone approssimazioni. Si può dimostrare che esse sono addirittura le migliori approssimazioni nel senso che $|q\alpha - p| > |q_0 \alpha - p_0|$ per ogni $0 < q < q_0$, se, e solo se p_0/q_0 è una delle frazioni ridotte di α.

3.3. Caratteristica dell'approssimazione e crescita dei quozienti parziali.

Un numero irrazionale α ammette approssimazioni razionali tanto migliori quanto piú velocemente crescono i suoi quozienti parziali a_i. Per esempio, se i

quozienti parziali sono limitati, la disequazione $|\alpha - (p/q)| < c/q^2$ in generale non ha soluzioni per c sufficientemente piccolo; se non sono limitati, essa ha sempre infinite soluzioni. Sono noti risultati ancora piú precisi.

Perciò è interessante la questione del comportamento dei quozienti parziali in funzione del numero α. Essi si annullano a un certo punto, se e solo se α è razionale. Essi sono periodici da un certo punto in poi, se e solo se α è un irrazionale quadratico, cioè $\alpha = \beta + \gamma\sqrt{\delta}$, con β, γ, δ razionali. Per molti irrazionali particolari sono note semplici formule per gli a_n; per esempio, $e = [2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, \ldots]$. Tuttavia l'autore non conosce un solo esempio di numero algebrico di grado superiore a due, per il quale sia noto se è limitata la sua successione di quozienti parziali. Dai tentativi fatti con i calcolatori con numeri quali $\sqrt[3]{2}$, o $\sqrt{2} + \sqrt{3}$ sembrerebbe che la risposta sia negativa.

Invece, risultati abbastanza definitivi sono stati ottenuti sul comportamento delle frazioni continue «in media», per quasi tutti i numeri α di $[0, 1]$ nel senso della misura di Lebesgue (nella quale ogni intervallo è misurabile e la sua misura è uguale alla sua lunghezza). Elenchiamone alcuni. L'insieme di tutti gli α con gli a_n limitati ha misura nulla. Se p_n/q_n è la n-esima frazione ridotta di α, per quasi tutti gli α (cioè per un insieme di misura 1) esiste il limite $\lim_{n\to\infty} \sqrt[n]{q_n}$ ed è uguale a $e^{\pi^2/12 \ln 2}$ (teorema di Lévy-Hinčin).

Sia $z_n(\alpha) = [0; a_n, a_{n+1}, a_{n+2}, \ldots]$ e sia $m_n(x)$ la misura dell'insieme $\{\alpha \mid z_n(\alpha) < x\}$; allora $\lim_{n\to\infty} m_n(x) = \ln(1+x)/\ln 2$ (congettura di Gauss dimostrata da Kuzmin e Lévy). Per quasi tutti gli α il limite $\lim_{n\to\infty} \sqrt[n]{a_1 \ldots a_n}$ esiste ed è uguale a $\prod_{r=1}^{\infty} \{1 + 1/r(r+2)\}^{\ln r/\ln 2} \simeq 2{,}6$. Invece, il limite $\frac{1}{n} \sum_{i=1}^{n} a_i$ non esiste; quasi ovunque infinite volte si ha $a_n > n \cdot \ln n$.

Non molto tempo fa si è scoperto che questi risultati sono direttamente interessanti per la cosmologia teorica. Precisamente, studiando le soluzioni delle equazioni della teoria generale della relatività di Einstein vicino alla singolarità iniziale – il momento della «grande esplosione» – sono stati scoperti dei regimi tipici, di carattere oscillatorio (regimi di Belinskij-Lifšic-Halatnikov). Ognuno di tali regimi è definito da un certo numero $\alpha \in [0, 1]$. Il regime che corrisponde a questo numero viene suddiviso in segmenti, che si chiamano ere, le quali sono in corrispondenza biunivoca con i quozienti parziali a_n del numero α. L'era corrispondente ad a_n viene suddivisa in a_n segmentini piú piccoli, le epoche. Internamente ad ogni epoca la metrica dello spazio localmente evolve approssimativamente come $t^{2p_1} dx^2 + t^{2p_2} dy^2 + t^{2p_3} dz^2$, dove $p_1 + p_2 + p_3 = p_1^2 + p_2^2 + p_3^2 = 1$ (soluzione di Kazner). Insieme con i tre numeri p_i sono definiti tre vettori, lungo uno dei quali lo spazio si estende, mentre lungo gli altri due si contrae. Cambiando epoca, l'asse dell'estensione e quello della contrazione piú lenta si scambiano, mentre l'asse della contrazione piú rapida resta quello precedente. Se cambia l'era, varia l'asse della contrazione piú rapida. La variazione degli esponenti (p_1, p_2, p_3) si scrive in funzione della frazione continua del numero α. Poiché, da un lato, tutta questa descrizione del regime è un'approssimazione matematica, e poiché, dall'altro lato, gli stessi modelli considerati possono cor-

rispondere solo approssimativamente alla realtà, ne segue che possono avere interesse solo le caratteristiche asintotiche dei regimi che hanno luogo per quasi tutti i valori di α; in particolare per la presenza delle perturbazioni che rendono incerti i valori teorici di α e di p_1, p_2, p_3.

4. *Aspetti algebrici della divisibilità.*

La teoria della divisibilità dei numeri interi serví come primo esempio non banale per una serie di costruzioni algebriche che hanno condotto, da una parte, all'aritmetica dei numeri algebrici e, dall'altra, alla moderna algebra commutativa con spiccata influenza della geometria (algebrica). Noi seguiremo l'evoluzione e la ramificazione degli aspetti della divisibilità in alcuni esempi fondamentali.

4.1. Gruppi abeliani.

La teoria delle congruenze forní consistenti esempi e metodi tecnici per studiare i gruppi abeliani astratti. Descriviamo i piú importanti teoremi strutturali.

Per i gruppi abeliani useremo la notazione additiva. Un sottoinsieme $G \subset A$ in un gruppo abeliano A si chiama sistema di generatori di A, se ogni elemento di A si può scrivere nella forma $\sum m_i g_i$, $m_i \in \mathbf{Z}$, $g_i \in G$. Se $(A_i)_{i \in I}$ è una famiglia di gruppi abeliani, si chiama somma diretta $\bigoplus_{i \in I} A_i$ l'insieme dei «vettori» $\{(a_i)_{i \in I} \mid a_i \in A_i$ e, per tutti gli i, ad eccezione di un numero finito, $a_i = 0\}$ in cui l'opposto e l'operazione di somma sono definiti coordinata per coordinata; $\bigoplus_{i \in I} A_i$ è un gruppo abeliano. Invece di $A \oplus \ldots \oplus A$ (m volte) scriveremo A^m.

Un gruppo A si dice ciclico se in esso vi è un sistema di generatori che consiste di un solo elemento g. È chiaro che, se tutti gli elementi di $\{mg \mid m \in \mathbf{Z}\}$ sono diversi, l'applicazione $\mathbf{Z} \to A : m \mapsto mg$ definisce un isomorfismo di $\mathbf{Z}$ su A. Se invece non tutti gli elementi mg sono diversi, allora esiste $d \neq 0$ con $dg = 0$ e tutti questi d si dividono per il piú piccolo d_0; da cui segue facilmente che l'applicazione $\mathbf{Z}/d_0\mathbf{Z} \to A$ definisce un isomorfismo di $\mathbf{Z}/d_0\mathbf{Z}$ su A. Pertanto otteniamo la classificazione dei gruppi ciclici a meno di un isomorfismo.

Un gruppo A si dice periodico, oppure gruppo di torsione, se per ogni elemento $a \in A$ esiste $m \in \mathbf{Z}$, $m \neq 0$, con $ma = 0$. Ogni gruppo periodico finito è isomorfo a una somma diretta di gruppi ciclici. Nel caso generale gli ordini di questi gruppi non sono definiti univocamente, ma vi sono due tipi di decomposizioni, le quali forniscono invarianti univoci. Prima di tutto vi è la decomposizione della forma $\bigoplus_{i=1}^{n} \mathbf{Z}/d_i\mathbf{Z}$, dove $d_1, \ldots, d_n$ è una successione di numeri naturali univocamente definita tali che d_i divide d_{i+1}. (Si chiamano divisori elementari del gruppo A). In secondo luogo vi è la decomposizione della forma $\bigoplus_{i,j} \mathbf{Z}/p_i^{r_{ij}}\mathbf{Z}$, dove p_i sono numeri primi e per ogni p_i è definita la successione degli esponenti $0 \leq r_{i1} \leq r_{i2} \leq \ldots \leq r_{ik_i}$ in modo che solo una quantità finita di numeri $p_i^{r_{ij}}$ sono diversi da 1. Qui la famiglia $(p_i^{r_{ij}})$ è definita univocamente.

Nelle dimostrazioni si usa essenzialmente la seguente circostanza: se $m, n \in \mathbf{Z}$ con $(m,n) = 1$, un gruppo abeliano finito A di ordine mn si scompone univocamente nella somma diretta dei suoi sottogruppi di ordine m e n. Infatti, se $mx + ny = 1$ $(x, y \in \mathbf{Z})$ e $a \in A$, si ha: $a = mxa + nya$, cosí che $A = mA \oplus nA = A_1 \oplus A_2$, dove $nA_1 = mA_2 = \{0\}$. In particolare $\mathbf{Z}/mn\mathbf{Z} = \mathbf{Z}/m\mathbf{Z} \oplus \mathbf{Z}/n\mathbf{Z}$.

Dal teorema della classificazione deriva il seguente criterio di ciclicità per un gruppo abeliano finito A: per ogni numero primo q la quantità di elementi di periodo q in A non supera q. Applicandolo al gruppo $(\mathbf{Z}/p\mathbf{Z})^*$, p-primo, troviamo che esso è ciclico, poiché il numero di soluzioni dell'equazione $z^q - 1 = 0$ nel corpo $\mathbf{Z}/p\mathbf{Z}$ non può essere superiore a q. Questo è il teorema sull'esistenza di un elemento primitivo nel gruppo delle classi dei resti modulo un numero primo. Esso è vero anche per i gruppi $(\mathbf{Z}/p^a\mathbf{Z})^*$ per $p \neq 2$, $a \geq 1$.

Un gruppo A non ha torsione se da $ma = 0$, dove $m \in \mathbf{Z}$ e $a \in A$, segue che $m = 0$, oppure $a = 0$. I gruppi abeliani senza torsione con un numero finito di generatori sono liberi, cioè isomorfi a $\mathbf{Z}^r$, $r \geq 0$. Il numero r si chiama rango; esso è definito univocamente e coincide con il minimo numero di generatori di A. La somma diretta di una famiglia infinita di gruppi $\mathbf{Z}$ si chiama gruppo abeliano libero di rango uguale alla potenza di questa famiglia.

Infine, ogni gruppo abeliano A con un numero finito di generatori è isomorfo alla somma diretta di un gruppo abeliano libero e di un gruppo abeliano finito. Il rango della componente libera e i divisori elementari del sottogruppo di torsione definiscono la classe di A a meno di un isomorfismo.

Useremo queste conoscenze per descrivere i risultati della teoria della divisibilità in anelli piú generali di $\mathbf{Z}$.

4.2. Proprietà moltiplicative degli anelli e dei corpi.

Sia A un anello commutativo con identità 1. Le nozioni seguenti generalizzano direttamente le definizioni del § 1 che si riferivano ad $A = \mathbf{Z}$.

Un elemento $a \in A$ divide l'elemento $b \in A$, se esiste $q \in A$ tale che $a = bq$. Un elemento $a \in A$ può dividere 1; allora si chiama unità dell'anello. Evidentemente a divide 1 se, e solo se, a è invertibile in A. L'insieme A^* delle unità è un gruppo commutativo. Abbiamo $\mathbf{Z}^* = \{\pm 1\}$; le unità nell'anello $\mathbf{Z}/m\mathbf{Z}$ sono state studiate nel § 2. Se A è un corpo e $A[x]$ è l'anello dei polinomi in una variabile su A, si ha $(A[x])^* = A^*$. La stessa cosa è vera per l'anello dei polinomi in piú variabili $A[x_1, \ldots, x_n]$.

In un anello A vi possono essere divisori dello zero non nulli; sono tali le classi $\bar{d}$, $1 < d < m$, $d \mid m$ nell'anello $\mathbf{Z}/m\mathbf{Z}$.

Se A non ha divisori dello zero, può essere immerso canonicamente in un corpo K, ogni elemento del quale si rappresenta nella forma ab^{-1}, $a, b \in A$, $b \neq 0$. Questo corpo si costruisce, a partire da A, quale corpo delle classi di coppie ordinate $\{(a,b) \mid b \neq 0\}$ con la relazione di equivalenza $(a,b) \sim (a',b')$, se $ab' = a'b$. La classe di (a,b) si indica con ab^{-1}; le operazioni si eseguono secondo le ordinarie formule scolastiche. K si chiama corpo dei quozienti dell'anello A.

Il corpo dei razionali $\mathbf{Q}$ è il corpo dei quozienti dell'anello dei numeri interi $\mathbf{Z}$. Il risultato fondamentale del § 1 si può riassumere cosí: il gruppo moltiplicativo $\mathbf{Q}^*$ è la somma diretta del gruppo $\{\pm 1\}$ e del gruppo abeliano libero di rango numerabile, liberamente generato dai numeri primi p_i. Un elemento $a \in \mathbf{Q}^*$ appartiene a $\mathbf{Z}$, se e solo se $a = \pm \prod p_i^{v_i}$ con $v_i \geq 0$.

Gli analoghi dei numeri primi in un anello commutativo A sono gli elementi non scomponibili, o primi. Un elemento $a \in A$ è primo, se non è l'identità e se ogni suo divisore è della forma u, oppure au, dove u è una unità in A. Ogni elemento au si dice associato ad a. Un anello A si dice anello a fattorizzazione unica, se esso non ha divisori dello zero e se il monoide $A - \{0\}$ (rispetto al prodotto), modulo A^*, è liberamente generato dalle classi degli elementi primi. In altre parole, scegliamo un elemento p in ogni classe di elementi associati. Allora in un anello a fattorizzazione unica ogni elemento si rappresenta in modo univoco nella forma $u \prod p_i^{v_i}$, $u \in A^*$, $v_i \geq 0$.

4.3. Anelli di polinomi e fattorizzazione unica.

Di tali anelli, oltre a $\mathbf{Z}$, fanno parte $k[x]$ e $k[x_1, \ldots, x_n]$, dove k è un corpo. Piú in generale, se A è un anello con fattorizzazione unica, ciò è vero anche per $A[x]$.

La dimostrazione nel caso di $k[x]$ e in una serie di altri casi, per esempio, per gli anelli di numeri interi gaussiani $\mathbf{Z}[i] = \{a + b\sqrt{-1} \mid a, b \in \mathbf{Z}\}$, si può ottenere generalizzando l'algoritmo della divisione con resto. La questione importante che bisogna risolvere sin dall'inizio sta nello stabilire come si deve confrontare il resto con il divisore, affinché questo resto esista e sia definito univocamente.

Nel caso dei polinomi in una variabile su di un corpo, l'uguaglianza $a = bq + r$ definisce q quale quoziente parziale e r quale resto nella divisione di a per b, se il grado di r è strettamente minore del grado di b ($r = 0$, se il grado di b è zero, cioè se b appartiene a k). Con questa definizione la divisione con resto è sempre possibile ed è univoca, e per l'anello $k[x]$ si possono ripetere le dimostrazioni del § 1 (sostituendo l'induzione sul valore del numero intero, con l'induzione sul grado del polinomio).

Gli elementi primi dell'anello $k[x]$ si dicono polinomi irriducibili sopra il corpo k. Per essi, con i metodi del § 1, si ottengono i seguenti risultati:

a) Il massimo comun divisore di due qualsiasi polinomi $a, b \in k[x]$ esiste, è definito univocamente a meno di un elemento di k^* e si rappresenta nella forma $au + bv$; $u, v \in k[x]$.

b) I polinomi irriducibili (a meno dell'associatività) sono infiniti. Se $k = \mathbf{C}$ (numeri complessi), tutti i polinomi irriducibili hanno la forma $a(x-b)$, $a \in \mathbf{C}^*$, $b \in \mathbf{C}$. Se $k = \mathbf{R}$, i polinomi irriducibili hanno la forma $a(x-b)$, $a \in \mathbf{R}^*$, $b \in \mathbf{R}$ e $ax^2 + bx + c$, $a \in \mathbf{R}^*$, $b, c \in \mathbf{R}$, $b^2 - 4ac < 0$. Se il corpo k è finito (per esempio, $k = \mathbf{Z}/p\mathbf{Z}$, p primo), vi è un numero finito di polinomi per ogni grado; perciò si hanno polinomi irriducibili di grado grande a piacere. Piú oltre preciseremo questo risultato.

Se tutti i polinomi irriducibili in $k[x]$ sono della forma $a(x-b)$, $a, b \in k$, il corpo k si dice algebricamente chiuso. Condizioni equivalenti: ogni polinomio ha una radice in k; ogni polinomio si scompone in fattori lineari in $k[x]$.

In questo caso, l'insieme dei polinomi irriducibili, a meno dell'associatività, è in corrispondenza biunivoca con gli elementi di k. Per $k = \mathbf{C}$, si ha il piano complesso. Considerando un polinomio $P(x)$ quale funzione su k, si ricava che la sua divisibilità per $x - a$ è equivalente al fatto che esso si annuli per $x = a$.

c) Analogamente alla definizione dell'anello $\mathbf{Z}/m\mathbf{Z}$, possiamo definire l'anello $k[x]/Pk[x]$, dove P è un qualsiasi polinomio (entrambe le costruzioni sono un caso particolare di anello quoziente rispetto a un ideale: si veda oltre). Se P è irriducibile, $k[x]/Pk[x]$ è un corpo K. Il suo grado su k, cioè la sua dimensione quale spazio lineare su k, è uguale al grado del polinomio P, l'applicazione canonica $k \to K : a \mapsto$ (classe di a modulo P) è una immersione di corpi. Il polinomio P, quale elemento di $K[x]$, non è piú irriducibile: da esso si estrae almeno un fattore lineare $x - \bar{x}$, dove $\bar{x} \in K$, $\bar{x} =$ classe di x mod P. Perciò la costruzione di K si chiama anche aggiunzione di una radice del polinomio P al corpo k.

Su ciò è basata la costruzione della chiusura algebrica di ogni corpo k, cioè di un corpo algebricamente chiuso $\bar{k}$ e dell'immersione $k \to \bar{k}$. Per costruire $\bar{k}$ aggiungiamo a k una radice per ogni polinomio irriducibile su k e indichiamo il corpo cosí ottenuto con k_1. Poi, a partire da k_1, costruiamo in modo analogo k_2; a partire da k_2 costruiamo il corpo k_3 e cosí via. L'unione $\bigcup_{i=0}^{\infty} k_i$ è la chiusura algebrica di k.

Due qualsiasi chiusure algebriche di un corpo k sono isomorfe (mediante un isomorfismo che è l'identità su k, ma in generale non sono identiche).

d) Sia $k = F_p = \mathbf{Z}/p\mathbf{Z}$, p numero primo. Si possono ottenere descrizioni assai dettagliate dei polinomi irriducibili P nell'anello $F_p[x]$ e dei corpi $K = F_p[x]/PF_p[x]$, detti corpi finiti di Galois.

Sia P un polinomio irriducibile di grado m. Allora il corpo K è di grado m su F_p e perciò consiste di p^m elementi. Le stesse considerazioni svolte per dimostrare il teorema di Fermat in $\mathbf{Z}/p\mathbf{Z}$ mostrano che $a^{p^m} = a$ per ogni $a \in K$. Applicando questa uguaglianza a $\bar{x} = x \bmod P(x)$, troviamo che il polinomio $x^{p^m} - x$ è divisibile per $P(x)$ in $F_p[x]$. Poiché $x^{p^m} - x$ è divisibile per $x^{p^d} - x$ per ogni $d \mid m$, si ricava che $x^{p^m} - x$ è divisibile pure per ogni polinomio irriducibile di grado $d \mid m$. Inversamente, se $x^{p^m} - x$ è divisibile per un polinomio irriducibile P di grado d, allora è $d \mid m$. Altrimenti, se r è il resto della divisione di m per d, ne segue che $x^{p^r} - x$ è divisibile per P, ma allora ciò significa che tutti gli elementi del corpo $F_p[x]/PF_p[x]$ sono radici di $x^{p^r} - x$, ciò che non può essere vero, poiché il numero di questi elementi è maggiore di p^r. Abbiamo cosí dimostrato che $x^{p^m} - x = \prod P(x)$, dove i P variano fra tutti i polinomi irriducibili con i gradi $d \mid m$ e con il primo coefficiente uguale a 1: il primo membro si divide per il secondo membro; inoltre il primo membro $x^{p^m} - x$ non ha fattori multipli perché, per esempio, la derivata di $x^{p^m} - x$ è uguale a -1. Indichiamo con

$i(d)$ il numero di polinomi irriducibili di grado d in $F_p[x]$. Da quanto precede, confrontando i gradi, si trova: $p^m = \sum_{d|m} d\, i(d)$. Ciò permette di trovare la funzione $i(m)$ nella forma $i(m) = \frac{1}{m} \sum_{d|m} \mu(d)\, d^{m/d}$, dove $\mu(d)$ è la funzione di Möbius: essa è uguale a 0, se d è divisibile per un quadrato >1, e $\mu(p_1 \cdots p_k) = (-1)^k$, se $p_1, \ldots, p_k$ sono numeri primi diversi tra loro.

Non è difficile verificare che $i(m) > 0$ per ogni m; in particolare, in $F_p[x]$ si hanno polinomi irriducibili di ogni grado m. I corpi $F_{p^m} = F_p[x]/PF_p[x]$ sono isomorfi per ogni scelta del polinomio irriducibile di grado m. Essi esauriscono tutti i possibili corpi finiti.

La formula per $i(m)$ risolve in un certo senso il problema della distribuzione dei polinomi irriducibili nell'anello $F_p[x]$. Questo problema è notevolmente piú elementare che quello della distribuzione dei numeri primi in $\mathbf{Z}$ e viene risolto evitando la funzione zeta. Però, già per le estensioni finite degli anelli $F_p[x]$ il problema corrispondente diventa molto profondo e viene risolto con mezzi di geometria algebrica. Esso è l'oggetto della teoria delle funzioni zeta delle curve algebriche e, piú in generale, delle varietà algebriche sui corpi finiti.

4.4. Esempi di anelli con fattorizzazione non unica.

Uno degli esempi piú semplici è fornito dall'anello $\mathbf{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} \mid a, b \in \mathbf{Z}\}$. In esso $21 = 3 \cdot 7 = (4+\sqrt{-5})(4-\sqrt{-5})$. Qui i fattori non sono scomponibili, né, a due a due, associati. L'assenza dell'univocità della fattorizzazione negli anelli $\mathbf{Z}[\sqrt[p]{1}]$, $p \geq 5$, è stata scoperta quale risultato dei tentativi di dimostrare la congettura di Fermat sull'assenza in $\mathbf{Z}$ di soluzioni dell'equazione $x^p + y^p + z^p = 0$, $xyz \neq 0$.

Lo studio profondo delle strutture moltiplicative negli anelli di numeri algebrici, iniziato da Kummer, Dirichlet ed altri, si è accompagnato con l'introduzione nella teoria della divisibilità dell'importantissimo concetto di ideale. Intuitivamente si può dire che, negli anelli piú complessi di $\mathbf{Z}$ le strutture additive e moltiplicative siano piú fortemente legate l'una all'altra, e nel concetto di ideale questo legame è espresso esplicitamente.

4.5. Ideali di anelli commutativi.

In un anello commutativo A si chiama ideale $\mathfrak{i}$ un qualsiasi sottoinsieme di A che: *a*) sia un sottogruppo additivo in A; *b*) sia chiuso rispetto al prodotto per elementi di A: per ogni $a \in A$, $b \in \mathfrak{i}$, si ha $ab \in \mathfrak{i}$.

Se $\mathfrak{i}$ è un ideale in A, imitando la costruzione dell'anello delle classi dei residui $\mathbf{Z}/m\mathbf{Z}$, possiamo costruire l'anello quoziente $A/\mathfrak{i}$. Gli elementi di $A/\mathfrak{i}$ sono i sottoinsiemi $a + \mathfrak{i} = \{a + b \mid b \in \mathfrak{i}\}$ in A. Invece di $a + \mathfrak{i}$ scriveremo $\bar{a}$, quando $\mathfrak{i}$ è fissato; a è un rappresentante della classe $\bar{a}$. Si chiama somma $\bar{a} + \bar{b}$ la classe $\overline{a+b}$ e prodotto $\bar{a}\bar{b}$ la classe $\overline{ab}$. L'indipendenza dalla scelta dei rappresentanti si verifica come nella teoria delle congruenze. Queste operazioni tra-

sformano $A/\mathfrak{i}$ in un anello; l'applicazione canonica $A \xrightarrow{\varphi} A/\mathfrak{i} : a \mapsto \bar{a}$ è un omomorfismo di anelli.

L'ideale $\mathfrak{i}$ viene ricostruito a partire da φ quale $\operatorname{Ker} \varphi$ (nucleo di φ), cioè quale controimmagine dello $\bar{0} \in A/\mathfrak{i}$: $\mathfrak{i} = \varphi^{-1}(\bar{0}) = \{a \in A \mid \varphi(a) = \bar{0}\}$. Se $\psi : A \to B$ è un qualsiasi omomorfismo di anelli commutativi, la controimmagine dello zero dell'anello B è un ideale in A, e se ψ è suriettiva, il diagramma $\psi : A \to B$ è isomorfo al diagramma $A \to A/\operatorname{Ker} \psi$. Pertanto la classificazione degli ideali in un anello è una variante strutturale della questione categoriale della classificazione degli omomorfismi di un anello A. Un ideale $\mathfrak{i}$ si dice primo, se $A/\mathfrak{i}$ non ha divisori dello zero. Condizione equivalente: da $ab \in \mathfrak{i}$ segue che $a \in \mathfrak{i}$, oppure $b \in \mathfrak{i}$.

Un ideale $\mathfrak{i} \subset A$ si dice principale, se esiste un elemento $a \in A$, generatore di $\mathfrak{i}$, tale che $\mathfrak{i} = aA = \{ab \mid b \in A\}$. La teoria delle congruenze è la teoria degli ideali principali nell'anello $\mathbf{Z}$ e degli anelli quozienti rispetto ad essi. Mostriamo che in $\mathbf{Z}$ tutti gli ideali sono principali. Infatti, sia $\mathfrak{i} \subset \mathbf{Z}$ un ideale. Se $\mathfrak{i} = \{0\}$, esso è principale. Se $\mathfrak{i} \neq \{0\}$, sia a il piú piccolo numero positivo contenuto in esso. Mostriamo che $\mathfrak{i} = a\mathbf{Z}$. Evidentemente $\mathfrak{i} \supseteq a\mathbf{Z}$, poiché $\mathfrak{i}$ è un sottogruppo additivo di $\mathbf{Z}$. Sia ora $b \in \mathfrak{i}$ un numero arbitrario. Dividendo b per a con resto, si ricava che $r = b - aq$ appartiene ad $\mathfrak{i}$ ($b \in \mathfrak{i}$; $a \in \mathfrak{i}$, perciò $aq \in \mathfrak{i}$). Ma $0 \leq r \leq a - 1$; poiché a è il piú piccolo numero positivo in $\mathfrak{i}$, deve essere $r = 0$ e $b = aq$, cosí che $\mathfrak{i} \subseteq a\mathbf{Z}$. Gli ideali primi in $\mathbf{Z}$ sono $\{0\}$ e $p\mathbf{Z}$, p numero primo.

Analogamente si dimostra che, se k è un corpo, in $k[x]$ tutti gli ideali sono principali (si considera nell'ideale un polinomio di grado minore e si dividono per esso con resto i polinomi dell'ideale). Gli ideali primi in $k[x]$ sono $\{0\}$ e $P(x)k[x]$, dove $P(x)$ è un polinomio irriducibile.

La semplicità dell'aritmetica negli anelli $\mathbf{Z}$ e $k[x]$ in misura notevole è dovuta all'assenza in essi d'ideali non principali.

Passando ad anelli piú grandi, quali $\mathbf{Z}[\sqrt{-5}]$, oppure $k[x_1, x_2]$, perdiamo questa proprietà; per esempio, l'ideale (x_1, x_2), che consiste dei polinomi in $k[x_1, x_2]$ con il termine noto nullo, non consiste di tutti i multipli di un solo polinomio.

La complessità della struttura degli ideali fu studiata classicamente in modo separato nella teoria dei numeri algebrici (per gli anelli del tipo $\mathbf{Z}[\sqrt{-5}]$, estensioni finite di $\mathbf{Z}$) e nell'algebra commutativa in geometria algebrica (per le estensioni finite degli anelli $k[x_1, \ldots, x_n]$). L'unione degli aspetti di teoria dei numeri e di quelli geometrici ha aperto la strada a molti successi ed a prospettive ancora piú ampie, poiché le intuizioni geometriche e di teoria dei numeri solitamente sono tra loro complementari. Qui ci limiteremo ad esporre alcuni elementi della teoria classica.

4.6. Caso della teoria dei numeri.

Un numero complesso $\alpha \in \mathbf{C}$ si dice algebrico, se esiste un polinomio $f(x)$ con coefficienti interi, del quale α sia radice. Il numero α si dice intero algebrico, se esiste un tale polinomio con il coefficiente direttivo uguale a 1.

Tutti i numeri algebrici formano un corpo $\bar{\mathbf{Q}} \subset \mathbf{C}$, la chiusura algebrica del corpo $\mathbf{Q}$. Esso non può essere generato da un numero finito di elementi su $\mathbf{Q}$. Quali approssimazioni si studiano corpi intermedi $\mathbf{Q} \subset K \subset \bar{\mathbf{Q}}$, di dimensione finita (in quanto spazi lineari) su $\mathbf{Q}$. Sia K uno di tali corpi; O l'anello di tutti i numeri algebrici interi in esso. Le differenze principali dell'aritmetica in O dall'aritmetica in $\mathbf{Z}$ sono legate alle unità non banali e agli ideali non principali.

4.7. **Unità** in O.

Prima di tutto in O vi possono essere le radici dell'identità, che formano un sottogruppo abeliano finito μ_K. Il gruppo quoziente O^*/μ_K è abeliano libero. Il suo rango si calcola in termini d'immersioni isomorfe di K nel corpo $\mathbf{C}$ dei numeri complessi. In tutto tali immersioni sono n = dimensione di K su $\mathbf{Q}$. Di esse siano r_1 le immersioni in $\mathbf{R}$ e $2r_2$ le immersioni in $\mathbf{C}$ (r_2 coppie di immersioni complesse coniugate). Allora il rango di O^*/μ_K è uguale a r_1+r_2-1 (teorema di Dirichlet).

4.8. **Ideali** in O.

L'insieme degli ideali non nulli in O può essere trasformato in un monoide rispetto alla legge di moltiplicazione $\mathfrak{ij} = \{\sum x_k y_k \mid x_k \in \mathfrak{i}, y_k \in \mathfrak{j}\}$. Questo monoide è commutativo e generato liberamente dagli ideali primi, cioè, per ogni $\mathfrak{i} \neq \{0\}$ ha luogo la decomposizione univoca $\mathfrak{i} = \mathfrak{p}_1^{v_1} \ldots \mathfrak{p}_k^{v_k}$, dove i $\mathfrak{p}_i$ sono ideali primi. D'altro lato, ogni elemento $a \in O$ genera l'ideale principale aO; gli ideali aO, bO coincidono se, e solo se, ab^{-1} è l'identità (oppure $a=b=0$). La decomposizione $aO = \mathfrak{p}_1^{v_1} \ldots \mathfrak{p}_k^{v_k}$ a livello di ideali è l'esatto analogo della decomposizione di un numero intero in fattori primi. Però, usare questa decomposizione per concrete indagini di teoria dei numeri nell'anello O (per esempio, per dimostrare la non-risolubilità delle equazioni) è notevolmente piú difficile. Oltre alla presenza delle unità, che, per $r_1+r_2>1$, fa sí che la ricostruzione di a, a partire dalla scomposizione di aO, abbia infiniti valori, non tutti gli ideali $\mathfrak{p}_1^{v_1} \ldots \mathfrak{p}_k^{v_k}$ sono principali. Una importante caratteristica della diversità è data dal gruppo delle classi di ideali del corpo K, che può essere definito quale gruppo quoziente di tutte le espressioni formali $\prod \mathfrak{p}_i^{v_i}$ (divisori, o frazioni di ideali) rispetto al sottogruppo generato dai divisori principali (cioè corrispondenti agli ideali aO). Il gruppo delle classi di ideali è sempre finito, ma raramente risulta banale.

4.9. **Linguaggio** della geometria algebrica.

In geometria algebrica ad ogni anello commutativo A si associa lo spazio topologico Spec A, spettro dell'anello A. I suoi punti sono gli ideali primi in A, mentre i suoi insiemi chiusi sono in corrispondenza biunivoca con gli ideali radicali di A: a un ideale $\mathfrak{i} \subset A$ corrisponde l'insieme chiuso $V(\mathfrak{i}) = \{\mathfrak{p} \in \operatorname{Spec} A \mid \mathfrak{p} \supset \mathfrak{i}\}$; un ideale $\mathfrak{i}$ è radicale, se da $x^n \in \mathfrak{i}$ segue che $x \in \mathfrak{i}$. D'altro lato, ogni elemento $f \in A$ viene considerato quale funzione sullo spazio Spec A, che assume valori in un insieme variabile: il valore di f nel punto $\mathfrak{p}$ è la classe dei residui di f mod $\mathfrak{p}$, considerata quale elemento del corpo dei quozienti dell'anello $A/\mathfrak{p}$. Queste funzioni sono continue nel senso che l'insieme degli zeri di ogni funzione è chiuso. La geometria dello spazio Spec A è un invariante molto fine di questo anello che rispecchia le sue strutture: sia quella additiva che quella moltiplicativa; in particolare, la relazione di divisibilità. Per meglio immaginare Spec A consideriamo il caso $A = \mathbf{C}[x_1, \ldots, x_n]$.

4.10. Ideali massimali.

In generale la topologia in Spec A è fortemente non separata; in particolare, in essa vi sono sottoinsiemi con un solo punto (o semplicemente punti) non chiusi. Un punto $\mathfrak{p} \in \operatorname{Spec} A$ è chiuso se, e solo se, l'ideale primo $\mathfrak{p} \subset A$ è massimale, cioè se ogni ideale intermedio tra $\mathfrak{p}$ e A coincide con $\mathfrak{p}$, o con A. In $\mathbf{C}[x_1, \ldots, x_n]$ tutti gli ideali massimali sono della forma $(x_1-\alpha_1, \ldots, x_n-\alpha_n) = \sum_{i=1}^{n} A(x_i-\alpha_i)$, dove $\alpha_i \in \mathbf{C}$ sono costanti arbitrarie. Perciò lo spettro massimale dell'anello A è lo spazio affine $\mathbf{C}^n = \{(\alpha_1, \ldots, \alpha_n) \mid \alpha_i \in \mathbf{C}\}$. (Un'affermazione analoga è vera sostituendo a $\mathbf{C}$ un qualunque corpo k algebricamente chiuso).

4.11. Insiemi chiusi.

In Spec $\mathbf{C}[x_1, \ldots, x_n]$ ogni insieme chiuso coincide con la chiusura dell'insieme dei punti chiusi in esso contenuti. A loro volta, questi insiemi chiusi di punti sono sottovarietà algebriche, cioè soluzioni di sistemi di equazioni della forma $\{F_i(x_1, \ldots, x_n) = 0\}$, dove F_i sono polinomi. Tutto ciò significa che l'aritmetica dell'anello dei polinomi coincide, o, piuttosto, è duale della geometria delle varietà algebriche nello spazio a n dimensioni.

5. *Aspetti topologici della divisibilità.*

Per m molto grande l'anello $\mathbf{Z}/m\mathbf{Z}$ è «vicino» all'anello $\mathbf{Z}$ nel senso che la somma e il prodotto delle classi dei resti dei numeri naturali non molto grandi ($<m/2$ per la somma; $<\sqrt{m}$ per il prodotto), si esegue come per questi stessi numeri naturali: il resto mod m è banale. Ma è piú comodo lavorare con l'anello $\mathbf{Z}/m\mathbf{Z}$ che non con $\mathbf{Z}$, poiché esso è finito e tutte le questioni, per principio, si risolvono mediante procedure finite. Su ciò è basata la dimostrazione della non-risolubilità di molte equazioni diofantee, quale conseguenza della non-risolubilità delle corrispondenti congruenze.

Per esempio, se $n \equiv 3 \bmod 4$, n non è somma dei quadrati di due numeri interi: i quadrati modulo 4 hanno resto 0, oppure 1 e la somma di due quadrati non può dare resto 3.

Entro certi limiti un'idea analoga, la sostituzione di un dominio infinito con uno grande ma finito, con successiva estensione periodica, viene spesso usata in fisica (per esempio, nella teoria del corpo rigido, nella teoria di Dirac delle radiazioni, ecc.). La questione su come si possa formulare quest'idea in un apparato matematico consistente non è banale e si risolve in modi diversi. Nella teoria dei numeri essa fu risolta costruendo i corpi di numeri p-adici, le loro estensioni, e cosí pure costruendo gli anelli di ideli e di adeli, nei quali topologia e aritmetica formano un tutto indivisibile.

5.1. Numeri p-adici.

Fissiamo un numero primo p. Introduciamo nell'anello dei numeri interi $\mathbf{Z}$ la topologia, nella quale un insieme $U \subset \mathbf{Z}$ è aperto se per ogni elemento $a \in U$ esiste un numero intero $k \geq 0$ tale che $a + p^k\mathbf{Z} \subset U$. È facile vedere che la somma, il prodotto e l'inversione di segno sono continui in questa topologia. In essa una successione $(a_k) \subset \mathbf{Z}$ converge ad $a \in \mathbf{Z}$, se $a - a_k$ è divisibile per una potenza di p che cresca indefinitamente con k. Per esempio, $n! \to 0$ per $n \to \infty$. Questa topologia si chiama p-adica.

Il gruppo $\mathbf{Z}$ non è completo in questa topologia: esistono successioni di Cauchy che non hanno limite in $\mathbf{Z}$ (per esempio, $a_k = 1 + p + \ldots + p^k$). Cosí come si costruiscono i numeri reali a partire dai razionali, definiamo l'anello $\mathbf{Z}_p$ quale completamento p-adico dell'anello $\mathbf{Z}$. Un suo elemento è la classe delle successioni di Cauchy nella topologia p-adica di $\mathbf{Z}$, modulo le successioni che convergono a zero. Somma e prodotto delle classi si definiscono coordinata per coordinata. L'applicazione $a \mapsto$ (classe $(a, \ldots, a, \ldots)$) definisce l'immersione $\mathbf{Z} \to \mathbf{Z}_p$. Nell'anello $\mathbf{Z}_p$ la topologia p-adica è definita esattamente come quella in $\mathbf{Z}$ (sostituendo $\mathbf{Z}$ con $\mathbf{Z}_p$). Rispetto a questa topologia $\mathbf{Z}_p$ è completo: ogni successione di Cauchy ha limite. Il sottoanello $\mathbf{Z}$ è denso in $\mathbf{Z}_p$: la sua chiusura coincide con $\mathbf{Z}_p$.

Gli elementi di $\mathbf{Z}_p$ si chiamano numeri interi p-adici. L'immersione $\mathbf{Z} \to \mathbf{Z}_p$ induce un isomorfismo degli anelli delle classi dei resti $\mathbf{Z}/p\mathbf{Z} \simeq \mathbf{Z}_p/p\mathbf{Z}_p$. Scegliamo in $\mathbf{Z}$ (oppure in $\mathbf{Z}_p$) un sistema arbitrario R di rappresentanti delle classi dei residui modulo p. Allora ogni numero intero p-adico si rappresenta in modo univoco mediante una serie $\sum_{\substack{i \geq 0 \\ a_i \in R}} a_i p^i$ convergente in $\mathbf{Z}_p$.

L'aritmetica in $\mathbf{Z}_p$ è semplice. I numeri non divisibili per p, e solo essi, sono invertibili. (Infatti, $1 + pa$, con $a \in \mathbf{Z}_p$, è invertibile: $(1 + pa)^{-1} = \sum_{i=0}^{\infty} (-1)^i a^i p^i$. I numeri interi b non divisibili per p sono invertibili: se $bx + py = 1$, $x, y \in \mathbf{Z}$, si ha: $bx = 1 - py$ invertibile e, quindi, b è un divisore dell'unità. Infine, ogni numero $c \in \mathbf{Z}_p$, non divisibile per p, si rappresenta nella forma $b(1 + pa)$, dove $b \in \mathbf{Z}$, $a \in \mathbf{Z}_p$). Ogni ideale in $\mathbf{Z}_p$ è principale; esso è generato da una certa potenza p^i, $i \geq 0$, oppure è $\{0\}$. Per ogni i si ha l'isomorfismo $\mathbf{Z}/p^i\mathbf{Z} \simeq \mathbf{Z}_p/p^i\mathbf{Z}_p$.

Il corpo dei quozienti $\mathbf{Q}_p$ dell'anello $\mathbf{Z}_p$ si chiama corpo dei numeri p-adici. Esso si potrebbe ottenere direttamente quale completamento del corpo $\mathbf{Q}$ nella

topologia p-adica, definita analogamente. È facile vedere che $\mathbf{Q}_p = \bigcup_{i=0}^{\infty} p^{-i}\mathbf{Z}_p$. Il vantaggio di $\mathbf{Z}_p$ e $\mathbf{Q}_p$ rispetto agli anelli $\mathbf{Z}/p^i\mathbf{Z}$ consiste nell'assenza di divisori dello zero e nella presenza di una topologia che permette di applicare mezzi d'indagine analitici. (Un esempio semplicissimo: definiamo il logaritmo p-adico mediante la serie $\log x = \sum_{i=0}^{\infty} (-1) \frac{i(x-1)^i}{i}$; essa converge per $x \in 1 + p\mathbf{Z}_p$ e l'applicazione $\log : (1 + p\mathbf{Z}_p)^* \to \mathbf{Z}_p$, $p > 2$, definisce un isomorfismo di gruppi topologici).

Questa topologia ha due caratteristiche che la distinguono dalla ordinaria topologia del continuo reale: essa è non-archimedea e totalmente sconnessa.

Dire che una topologia non è archimedea significa che, se $\alpha \in \mathbf{Z}_p$ è un elemento piccolo nel senso p-adico, ogni suo multiplo intero $m\alpha$ è altrettanto piccolo (cioè è divisibile per una potenza di p non inferiore). Piú esattamente, l'anello $\mathbf{Z}_p$ (o il corpo $\mathbf{Q}_p$) è dotato di una ultrametrica. Ponendo, per $\alpha \in p^k\mathbf{Z}_p^*$, $|\alpha|_p = p^{-k}$, si ricava che l'applicazione $|\ |_p : \mathbf{Z}_p \to \mathbf{R}$ è una norma p-adica con le proprietà seguenti: *a*) $|\alpha|_p = 0$, se, e solo se, $\alpha = 0$; *b*) $|\alpha\beta|_p = |\alpha|_p |\beta|_p$; *c*) $|\alpha + \beta|_p \leq \max(|\alpha|_p, |\beta|_p)$ (viene cosí rafforzata la disuguaglianza triangolare; da qui il nome di ultrametrica per $|\ |_p$; per l'ordinario valore assoluto, o per il modulo dei numeri complessi, si ha $|\alpha + \beta| \leq |\alpha| + |\beta|$). Ciò semplifica molti aspetti dell'analisi p-adica. Per esempio, la serie $\sum \alpha_i$, $\alpha_i \in \mathbf{Q}_p$, converge se, e solo se, $|\alpha_i|_p \to 0$.

Il fatto che le topologie in $\mathbf{Q}_p$ e $\mathbf{Z}_p$ siano totalmente sconnesse significa che vi è la possibilità di rappresentare ogni sottoinsieme chiuso come unione di due sottoinsiemi chiusi disgiunti. In altre parole, le componenti connesse massimali in $\mathbf{Q}_p$ e $\mathbf{Z}_p$ sono i singoli punti. Ciò complica la teoria delle funzioni analitiche p-adiche, poiché rende inapplicabile l'ordinaria definizione del prolungamento analitico: il teorema dell'unicità del prolungamento di un germe di una funzione analitica secondo la definizione ingenua è falso senza speranza. Per questo una analisi p-adica globale è stata formulata solo negli anni recenti, quando la geometria algebrica e analitica hanno dato modelli di teorie nelle quali gli aspetti geometrici e analitici erano strettamente fusi sin dall'inizio.

5.2. Estensioni finite dei corpi p-adici.

Con la chiusura di $\mathbf{Q}$ nella topologia ordinaria otteniamo i numeri reali $\mathbf{R}$. Questo corpo non è algebricamente chiuso, ma la sua chiusura algebrica $\mathbf{C}$ è di grado 2 su esso. Il corpo $\mathbf{C}$ dei numeri complessi è contemporaneamente algebricamente chiuso e topologicamente completo.

La situazione con il corpo $\mathbf{Q}_p$ è notevolmente piú complessa. Esso pure non è, come $\mathbf{R}$, algebricamente chiuso, ma la sua chiusura algebrica $\bar{\mathbf{Q}}_p$ ha su $\mathbf{Q}_p$ grado infinito. Di piú, $\bar{\mathbf{Q}}_p$ non è topologicamente completo e il suo completamento $\hat{\bar{\mathbf{Q}}}_p$ è ancora una estensione a infinite dimensioni su $\bar{\mathbf{Q}}_p$. Il corpo $\hat{\bar{\mathbf{Q}}}_p = \mathbf{C}_p$ è l'analogo p-adico dei numeri complessi; esso è algebricamente chiuso e topologicamente completo. Ma esso è molto piú lontano da $\mathbf{Q}_p$ che non $\mathbf{C}$

da **R**. Inoltre in esso vi è una brutta topologia (per esempio, non localmente compatto) ed è ancora molto poco studiato. Lo studio del gruppo di Galois del corpo $\bar{\mathbf{Q}}_p$ su $\mathbf{Q}_p$ costituisce l'oggetto della teoria locale dei corpi di classi. Ogni estensione algebrica finita del campo $\mathbf{Q}_p$ si può ottenere come completamento di un'estensione finita K del campo **Q** rispetto ad una topologia $\mathfrak{g}$-adica, dove $\mathfrak{g}$ è l'ideale primo, divisore di p nell'anello dei numeri interi del campo K.

5.3. Adeli e ideli.

La topologia reale di **Q** e le sue topologie p-adiche esauriscono tutte le topologie che possono essere date mediante norme con la disuguaglianza triangolare. Per molti aspetti è comodo considerare tutte queste topologie su uno stesso piano in un arco di questioni possibilmente piú esteso. Chiamiamo ognuna delle immersioni $\mathbf{Q}\to\mathbf{Q}_p$, oppure $\mathbf{Q}\to\mathbf{R}$ del corpo **Q** in uno dei suoi completamenti *punto* di questo corpo; i punti $\mathbf{Q}\to\mathbf{Q}_p$ si dicono finiti, mentre il punto $\mathbf{Q}\to\mathbf{R}$ si dice infinito. (Questo linguaggio classico concorda con la terminologia della geometria algebrica, nella quale i punti $\mathbf{Q}\to\mathbf{Q}_p$ corrispondono agli elementi massimali in Spec **Z**). L'oggetto piú grande, che riunisce tutte queste immersioni, è l'omomorfismo $\mathbf{Q}\to\prod_v \mathbf{Q}_v$, dove v si estende a tutti i punti di **Q** sia finiti che infiniti. Gli elementi di $\prod \mathbf{Q}_v$ sono successioni infinite $(\alpha_\infty, \alpha_2, \alpha_3, \ldots)$, dove $\alpha_\infty \in \mathbf{R}$, $\alpha_p \in \mathbf{Q}_p$ (al numero $\alpha \in \mathbf{Q}$ è associata la successione $(\alpha, \alpha, \ldots)$) e dove ogni coordinata di questo vettore viene considerata quale elemento del corrispondente completamento.

L'anello $\prod_v \mathbf{Q}_v$ è molto grande. In esso gli elementi $(\alpha, \alpha, \ldots)$, $\alpha \in \mathbf{Q}$, soddisfano una limitazione ulteriore: tutte le loro coordinate, ad eccezione di un numero finito, sono numeri interi nel corrispondente completamento (fanno eccezione i divisori del denominatore di α e, eventualmente, ∞). L'insieme di tutti gli elementi di $\prod_v \mathbf{Q}_v$ che soddisfano questa condizione si chiama anello degli adeli del corpo **Q** (questo insieme è chiuso rispetto alla somma e al prodotto e contiene 0 e 1). L'insieme di tutti gli elementi invertibili di questo anello si chiama gruppo degli ideli del corpo **Q**.

Queste definizioni si estendono senza molte variazioni ad ogni estensione finita K del corpo **Q**. I suoi punti finiti sono le immersioni $K\to K_{\mathfrak{g}}$, dove $\mathfrak{g}$ sono gli ideali primi dell'anello degli interi di K, mentre $K_{\mathfrak{g}}$ sono i completamenti corrispondenti. I suoi punti infiniti sono l'immersione di K in **R**, o le coppie di immersioni complesse coniugate di K in **C**; **R** o **C** sono i completamenti corrispondenti. L'anello A_K degli adeli è il sottoanello di $\prod_{\mathfrak{v}} K_{\mathfrak{v}}$ costituito dai vettori, le cui coordinate sono quasi ovunque intere; il gruppo degli ideli di K è il gruppo degli elementi invertibili di A_K: i suoi elementi hanno tutte le coordinate non nulle, e queste sono quasi ovunque unità $\mathfrak{g}$-adiche.

Il linguaggio degli adeli e degli ideli è largamente usato nelle indagini contemporanee nella teoria dei numeri algebrici, nella teoria dei corpi di classi, nell'aritmetica delle varietà algebriche. [JU. I. M.].

Fin dai tempi della filosofia naturale dell'antica Grecia s'è posto il problema della divisibilità della **materia**, dello **spazio/tempo** e s'è rivelata l'opposizione **continuo/discreto**, alla quale si riconducono in particolare le teorie atomistiche (cfr. **atomo e molecola** e, per le aporie relative alla divisibilità dello spazio ed al moto, **dialettica**). In questo articolo tuttavia si trattano precipuamente gli aspetti teoretico-numerici e algebrici della divisibilità.

I principali concetti di essa vengono presentati in rapporto ai numeri naturali; in **applicazioni** si illustrano situazioni piú generali ove ha senso l'operazione di divisibilità. Ma già l'ambito ristretto dei numeri è sufficiente a presentare concetti e strumenti – come il massimo comun divisore o l'**algoritmo** d'Euclide – di grande forza penetrativa. Ed i problemi della **teoria** dei numeri (cfr. **numero**) sono tra i piú sottili ed intricati della matematica. **L'approssimazione** dei reali mediante razionali è profondamente legata, oltre che ad esigenze pratiche immediate (cfr. **calcolo**), ad una vasta gamma di fenomeni, come la cosmologia teorica (cfr. **stabilità/instabilità, cosmologie**) o la teoria dei suoni musicali (cfr. **suono/rumore**) e l'**armonia.**

Gli aspetti algebrici della divisibilità mettono in rilievo l'importanza delle **strutture matematiche**. Nell'ambito generale della teoria degli anelli è possibile cogliere i legami fra la teoria dei numeri e la geometria algebrica (cfr. **geometria/topologia, curve e superfici**), evidenziando il ruolo della **dualità**, e la distinzione fondamentale **razionale/algebrico/trascendente.**

Dualità

Scopo principale di questo articolo è la descrizione dei differenti impieghi della parola 'dualità' in quanto termine matematico. La dualità, e i termini analoghi (coniugazione, reciprocità) figurano nella denominazione sia di costruzioni generali (dualità degli spazi lineari, funtori aggiunti) che di teoremi particolari (dualità di Poincaré, legge di reciprocità di Artin). In tutte le varie utilizzazioni della parola è contenuta l'idea di simmetria bilaterale di un oggetto, una costruzione o una teoria matematica. Tale simmetria può non essere completa: gli oggetti legati dalla dualità possono avere natura differente.

L'idea di dualità ha un carattere estremamente generale. È naturale che essa trovi una sua espressione anche in altri campi della ricerca scientifica. Fra le correlazioni piú vicine alla dualità matematica si trovano, ad esempio, il principio delle classificazioni dicotomiche, le «opposizioni» dello strutturalismo contemporaneo (linguistica, etnografia), la «complementarità» nella fisica quantistica, inclusa l'interpretazione piú ampia che ne dà Niels Bohr. Una esposizione puramente matematica della teoria della dualità, senza alcun legame con tali concezioni, sarebbe unilaterale. Pertanto l'introduzione sarà dedicata a questi aspetti meno formali della dualità.

Dualità fenomenologiche. In varie discipline si possono individuare numerose contrapposizioni a coppie nel sistema dei concetti fondamentali direttamente connessi con il materiale osservato. Cosí, dai dati di sistemi mitologici differenti si ricostruiscono le opposizioni fondamentali della concezione arcaica del mondo: caos/ordine (nei miti cosmologici), terrestre/celeste, sinistro/destro, rosso/nero, maschile/femminile. Esse si fondano sull'esperienza diretta, anche se subiscono una complessa simbolizzazione nel sistema. A questa stessa classe si possono ricondurre quelle coppie di categorie della scienza contemporanea come fenotipo/genotipo in biologia o particella/antiparticella in fisica. La scoperta di queste coppie spesso è accompagnata anche dalla scoperta di una loro disparità di ruoli (rottura della simmetria). Nel nostro mondo predomina nettamente la materia, e non l'antimateria. Negli organismi viventi viene usata e prodotta solo una delle due forme specularmente opposte di sostanze otticamente attive (Pasteur): ad esempio nell'uomo la forma destrogira del glucosio, e non quella levogira del fruttosio. Le opposizioni, che regolano importanti aspetti della vita sociale e della cultura, non sono simmetriche: uno degli elementi della coppia viene considerato «proprio», «buono», «normativo», in contrapposizione a «estraneo» e «cattivo». Si vedano le tracce dell'antica relazione della destra con il buono, con il giusto, e della sinistra con il cattivo, con l'impuro, ecc., nella semantica delle parole moderne: in inglese *right*, *sinister*, in francese *droit*, in russo правый.

Dicotomie e contrapposizioni. Ad un livello superiore di astrazione rispetto alle coppie considerate in precedenza appartengono le opposizioni «presenza/

assenza di contrassegni». Il principio della classificazione dicotomica di alcune classi di oggetti consiste nel fatto che *a*) si sceglie una certa famiglia di contrassegni, e *b*) ogni oggetto viene caratterizzato in relazione a tale famiglia da una successione, ad esempio, di zeri (assenza del dato contrassegno) ed uni (sua presenza). Cosí Jakobson, Fant, Halle hanno proposto una classificazione dicotomica di tutti i possibili suoni del discorso nelle diverse lingue del mondo, fondata su dodici coppie di contrassegni differenziali, tra i quali vocalico/non-vocalico, consonantico/non-consonantico, nasale/orale, ecc. Secondo questi contrassegni *r* è vocalico/consonantico/orale (110), e й (i-breve) è non-vocalico/non-consonantico/orale (000). La classificazione in generale è uno dei piú semplici e piú importanti strumenti dell'indagine scientifica. La classificazione dicotomica dal punto di vista della matematica conduce ad una scoperta importante: la crescita logaritmica, cioè molto lenta, della quantità d'informazione in dipendenza dalla molteplicità degli oggetti studiati. Sei risposte sí/no permettono di descrivere univocamente un qualsiasi oggetto (suono del discorso) scelto fra 64, il che basta per la maggior parte delle lingue del mondo; dodici segni, in linea di principio, possono bastare per la descrizione di $2^{12} = 4096$ suoni. Una risposta sí/no porta una unità d'informazione. L'organizzazione del materiale secondo l'antico principio della dicotomia conduce a numerare ogni oggetto con un numero binario, come nella notazione posizionale molto diffusa nei calcolatori contemporanei. Una buona classificazione dicotomica, evidentemente, si differenzia da un'arbitraria numerazione binaria per il fatto che le posizioni degli zeri/uni non sono arbitrarie, ma corrispondono a contrassegni rilevanti degli oggetti. Ogni posizione definisce una famiglia di oggetti contrapposti a due a due, che differiscono solo per il contrassegno ad essa corrispondente.

Tutte queste caratteristiche della dicotomia, che spiegano la sua fondamentale importanza, raramente s'incontrano allo stato puro. La discretizzazione della scala continua di gradazioni di un contrassegno in due valori può risultare troppo rozza per altri scopi. Non tutte le 2^n combinazioni di n contrassegni possono risultare possibili o equiprobabili: cosí la vocalicità del suono si combina soprattutto con la sua non-consonanticità, e viceversa, in accordo con la grammatica elementare. Queste correlazioni o regole di interdizione definiscono una certa struttura nell'insieme dei contrassegni, che spesso è piú importante della struttura dell'insieme degli oggetti: le leggi scientifiche si possono formulare in termini di proprietà, ma non di cose.

Complementarità. Al livello successivo l'idea di dualità si traduce nell'individuazione di alcune coppie di categorie di carattere metodologico. In termini di queste categorie si può descrivere il metodo di un'indagine concreta, il procedimento di elaborazione del materiale e di rappresentazione dei risultati. Inoltre in ogni indagine concreta domina in genere soltanto uno dei membri della coppia.

Categorie di questo genere sono: analisi/sintesi, soggetto/oggetto, causa/fine (in biologia: un organo esiste perché..., oppure affinché...), immagine/sfondo, sincronia/diacronia (in linguistica), continuo/discreto, ecc. Molte di queste cop-

pie manifestano gli aspetti caratteristici della complementarità. Essa si manifesta nel fatto che la descrizione in termini di uno qualsiasi dei membri della coppia è incompleta, mentre le descrizioni simultanee in termini di entrambi i membri risultano incompatibili. Nello stesso tempo nessun elemento della coppia è semplicemente la negazione dell'altro nel senso classico della parola.

Il termine stesso 'complementarità' è stato introdotto da Niels Bohr all'epoca della fondazione della meccanica quantistica per caratterizzare opposizioni quali onda/particella. Non è questo il luogo per entrare nei particolari del significato classico di tali termini. È sufficiente dire che la luce, descrivibile classicamente con equazioni d'onda, manifesta la sua natura corpuscolare ad esempio nell'effetto fotoelettrico, dove ogni atto di assorbimento di luce di frequenza ν da parte di un atomo è accompagnato dal trasferimento di un'energia $h\nu$ (o, con piccola probabilità, di un multiplo intero di $h\nu$, ma mai di, ad esempio, $h\nu/2$). Queste porzioni sono i fotoni. Viceversa, come è stato scoperto in seguito, una particella come l'elettrone, che senza dubbio dal punto di vista classico è una particella e non un'onda, nella diffusione in un reticolo cristallino manifesta la sua natura ondulatoria. Questo dualismo si concretizza oggi in apparecchi scientifici di uso corrente: l'astronomo che, con esposizione di molte ore, raccoglie i fotoni su una lastra fotografica, sfrutta i «corpuscoli di luce», mentre il biologo che lavora col microscopio elettronico riesce ad ottenere un ingrandimento impossibile nella banda ottica grazie alla piccolezza della lunghezza d'onda associata all'elettrone.

Per Bohr la complementarità degli aspetti ondulatorio/corpuscolare consisteva, in particolare, nell'impossibilità che entrambi questi aspetti si manifestassero simultaneamente in un unico esperimento. Solo l'interazione della radiazione o del fascio di elettroni con un sistema macroscopico di tipo speciale (lo strumento) conduce al fenomeno osservato, che noi classifichiamo nei casi estremi come ondulatorio o corpuscolare. Nei casi intermedi entrambi i modi di descrizione classici risultano solo limitatamente applicabili. Inoltre alcune coppie di grandezze classiche (come posizione/impulso di una particella) vengono a trovarsi, in meccanica quantistica, in rapporto di complementarità, cessando di essere contemporaneamente misurabili con un arbitrario grado di precisione. Nel § 3 dell'articolo si descriveranno i modelli matematici di tali coppie di variabili coniugate, e si darà la formulazione precisa del «principio d'indeterminazione» di Heisenberg, che stima il prodursi di una dispersione attorno al valor medio di due osservabili che non commutano. A questo principio si è attribuita un'importanza particolare nei primi anni della fondazione della meccanica quantistica. Oggi noi comprendiamo che esso è solo una semplice conseguenza dei suoi postulati fondamentali, dalla quale questi stessi postulati non possono in alcun modo essere dedotti.

Senza riferirsi a modelli teorici particolari, Bohr mise in luce un aspetto qualitativo della complementarità quantistica: l'influenza inevitabile e non prevedibile dell'osservazione sul sistema osservato. In tale formulazione la complementarità ha un'implicazione filosofica, presentando sotto una nuova luce il dualismo soggetto/oggetto che pervade tutta la scienza. In tutti i procedimenti delle scienze

naturali il soggetto osservante analizza alcune caratteristiche del sistema osservato, che egli si rappresenta come completamente isolato attraverso l'esclusione delle azioni estranee. In effetti all'atto stesso della misura il sistema cessa di essere isolato, diventando parte di un unico complesso osservatore+osservato. Perciò si analizzano in realtà alcuni parametri di questo sistema complessivo. È necessario ancora sottolineare che in meccanica quantistica l'unificazione di questi sistemi comporta dei legami forti e non classici tra le parti.

Dualità, matematica e linguaggio. Anche i primi esempi di dualità matematica hanno avuto origine da «materiale osservabile»: numeri e figure. Se ne porteranno qui due esempi: la costruzione del piano proiettivo e la legge di reciprocità quadratica.

Piano proiettivo. Sul consueto piano euclideo per due qualsiasi punti passa esattamente una retta, ma non è detto che due rette arbitrarie si intersechino esattamente in un punto: le coppie di rette parallele sono un'eccezione. Poncelet e Gergonne (intorno al 1820) dimostrarono che di questa asimmetria ci si può liberare, estendendo il piano euclideo a quello proiettivo. Il piano proiettivo si ottiene da quello euclideo aggiungendovi una «retta all'infinito», i cui punti siano in corrispondenza biunivoca con i fasci di rette parallele del piano euclideo. Ogni retta L interseca la retta all'infinito nel punto che corrisponde al fascio delle rette parallele ad L.

L'intero insieme dei punti e delle rette del piano proiettivo soddisfa alle seguenti condizioni:

1) Per due punti arbitrari distinti passa esattamente una retta.
2) Due rette arbitrarie distinte si intersecano esattamente in un punto.

Una serie di altri assiomi della geometria piana assume un aspetto simmetrico rispetto allo scambio retta/punto con la conservazione della relazione «la retta ed il punto sono incidenti» (che è l'equivalente simmetrico delle relazioni «il punto giace sulla retta» e «la retta passa per il punto»). Perciò tale sostituzione permette di raddoppiare automaticamente il numero di teoremi sul piano proiettivo. La peculiarità di questo esempio consiste nell'introduzione del nuovo concetto di punti «ideali» o «all'infinito» del piano euclideo, le proprietà dei quali sono dettate proprio dalla condizione che la simmetria incompleta di un oggetto iniziale acquisti completezza.

La dualità proiettiva è stata una delle fonti della piú generale e profonda teoria della dualità degli spazi lineari, per i cui aspetti algebrici cfr. § 2.1.

Legge di reciprocità quadratica. Siano n ed m due numeri interi. Un numero m si chiama residuo quadratico modulo n se esiste un intero x tale che m ed x^2 divisi per n dànno lo stesso resto. Nel caso contrario m viene detto non-residuo quadratico. Ad esempio 3 è un non-residuo quadratico modulo 5 e 7, ma è un residuo quadratico modulo 11 (5^2 diviso 11 dà resto 3).

Si definisce il simbolo di Legendre $\left(\frac{m}{n}\right)$, per due numeri interi m ed n primi

tra loro, mediante la condizione $\left(\frac{m}{n}\right)=1$ se m è residuo quadratico modulo n; $\left(\frac{m}{n}\right)=-1$ in caso contrario.

La parte piú importante e di piú semplice formulazione del teorema detto legge di reciprocità quadratica afferma che se p e q sono due differenti numeri primi dispari, allora $\left(\frac{p}{q}\right)\left(\frac{q}{p}\right)=(-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}$. Per esempio, $\left(\frac{3}{7}\right)\left(\frac{7}{3}\right)=(-1)\cdot(1)=$ $=(-1)^{1\cdot 3}$.

Che questa simmetria sia inattesa dipende dal fatto che nella definizione del simbolo $\left(\frac{p}{q}\right)$ i numeri p e q entrano in modo molto asimmetrico: q come divisore e p come dividendo.

Come fatto sperimentale la legge di reciprocità avrebbe potuto essere osservata fin dai tempi dell'antichità (fornendo un bellissimo esempio di magia pitagorica dei numeri). In realtà fu scoperta da Eulero, e la prima (incompleta) dimostrazione fu data da Legendre. Gauss propose otto differenti dimostrazioni della legge di reciprocità quadratica. Esse si basano su principî diversi e mettono in luce aspetti diversi di questa notevole simmetria. Dal punto di vista contemporaneo essa è risultata inclusa nella profonda teoria dei corpi di classi, in cui hanno importanza alcune dualità generali – la dualità di Galois (estensioni di campi/gruppi); la dualità dei gruppi abeliani (gruppi/caratteri); la dualità di Fourier – e le loro concrete realizzazioni aritmetiche.

Proprio queste significative osservazioni matematiche furono storicamente il fondamento dell'individuazione del concetto generale di dualità. Nel § 1 saranno esposti alcuni esempi fondamentali che illustrano la struttura algebrica e le applicazioni della dualità. Il § 2 è dedicato al ruolo del linguaggio categoriale, che permette di formalizzare ed unificare sotto dualità concrete una serie di osservazioni in precedenza eterogenee. Infine il § 3 è dedicato alla relazione fra dualità e continuità.

Va sottolineato che, nonostante il continuo ampliarsi dello spettro delle costruzioni matematiche concrete, in matematica resta utile la stessa idea indefinita di dualità, potenzialmente aperta e non ridotta a nessuno dei contesti in cui la parola 'dualità' sia usata come termine tecnico.

A maggior ragione ciò riguarda la semantica delle dualità matematiche, quando esse appaiono nell'apparato delle scienze applicate.

Per esempio si può notare che nella maggior parte delle dualità matematiche si considerano versioni concrete della opposizione generale locale/globale nei suoi diversi aspetti. Esempio tipico: un vettore nello spazio (locale) / un funzionale sullo spazio (globale). Oltre a ciò, le piú importanti costruzioni della dualità ammettono come sovrastruttura alcuni nuovi oggetti che sintetizzano i membri della coppia in un'unica struttura spesso simmetrica rispetto a questi membri. Allora il confronto dei singoli membri della coppia con il locale o con il globale perde il suo significato assoluto: ciascuno di essi può essere locale da un certo punto di vista e globale da un altro punto di vista. Proprio questo è il meccanismo

di superamento della contraddizione onda/particella in meccanica quantistica.

Questo esempio è caratteristico: esso mostra come la matematica svolga un ruolo di effettiva mediazione nella descrizione della sintesi dei membri complementari di una opposizione. La semantica della matematica, come linguaggio, è libera; uno «spazio lineare» dal punto di vista di una interpretazione «realistica» è un recipiente vuoto, pronto per essere riempito di significati diversi. Uno spazio lineare e lo spazio lineare ad esso coniugato formano una coppia universale di concetti, la simmetria della cui descrizione matematica libera il cervello dalle restrizioni accumulate dalla semantica del linguaggio comune. Le possibilità di quest'ultimo, in tale campo, sono estremamente limitate e non offrono una valida alternativa (cfr. anche i modelli matematici e filosofici delle opposizioni causa/fine, casuale/deterministico).

D'altra parte il linguaggio naturale possiede senza dubbio alcune qualità complementari (le quali hanno permesso, ad esempio, di descrivere proprio il concetto di complementarità senza essere legati a sue realizzazioni troppo rigide).

Il pensiero mitologico, poetico, umanistico (ed a livelli diversi anche delle scienze naturali) trova un altro sbocco risolvendo i suoi problemi di sintesi mediante la complementarità: esso edifica costruzioni linguistiche lontane dalla precisione, ambigue, irritanti, ma che stimolano l'immaginazione. La correlazione, già rilevata da Schleicher in molti linguaggi indoeuropei, dell'idea di 'due' con l'idea di 'dubbio', 'timore' (o 'duplicazione') (si veda il tedesco *Zwei/Zweifeln*, il greco δοιός/δοιή, il latino *duo/dubius*) è riscontrabile a tutti i livelli dell'attività linguistica sociale, fino alla comparsa della «doppiezza» come tema della letteratura mondiale (dai «romanzi del terrore» fino a Dostoevskij). La naturale reazione positivistica all'abbondanza di testi ambigui nella letteratura al confine fra scienza e filosofia della natura era inevitabile. L'espressione piú concisa di tale reazione è nella famosa tesi di Wittgenstein «ciò che in generale può essere detto può essere detto chiaramente, e di ciò di cui non si è in grado di parlare occorre tacere». Secondo l'ironica classificazione di Bohr questo detto esprime senza dubbio una profonda verità: esso è vero assieme alla sua negazione, che è anch'essa una profonda verità. Tuttavia val la pena di pensare agli enormi filoni della cultura e dell'attività umana sui quali incombe il divieto dell'imperativo di Wittgenstein inteso letteralmente. Il medico è disposto ad apprezzare grandemente la chiarezza di giudizio del fisiologo, ma egli stesso è costretto ad elaborare troppi fatti in troppo poco tempo ed a parlare, non riuscendo ad essere chiaro.

1. *Esempi fondamentali: discreti e a dimensione finita.*

1.1. Dualità booleana.

Formalismo. Si consideri un insieme B, nel quale siano dati due elementi specificati $0,1\in B$, due leggi di composizione binarie $\wedge$ (intersezione o moltiplicazione) e $\vee$ (unione o somma), e un'operazione $'$ (complemento). La strut-

tura $\mathbf{B}=\{B;\, 0, 1;\, \wedge, \vee;\, '\}$ si chiama algebra booleana se sono soddisfatti i seguenti assiomi:

1) Per tutti gli $a \in B$ si ha $(a')'=a$, $0'=1$, $1'=0$.
2) Le operazioni $\wedge$, $\vee$ sono associative e commutative, cioè per ogni $a, b, c \in B$

$$a \wedge b = b \wedge a \qquad (a \wedge b) \wedge c = a \wedge (b \wedge c)$$
$$a \vee b = b \vee a \qquad a \vee (b \vee c) = (a \vee b) \vee c.$$

3) Le operazioni $\wedge$, $\vee$ sono distributive l'una rispetto all'altra, cioè per ogni $a, b, c \in B$

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c) \qquad a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c).$$

4) Le operazioni $\wedge$, $\vee$ sono duali rispetto a $'$, cioè per ogni $a, b \in B$

$$(a \vee b)' = a' \wedge b' \qquad (a \wedge b)' = a' \vee b'.$$

5) Per ogni $a \in B$ si ha $a \vee a = a \wedge a = a$, $1 \wedge a = a$ e $0 \vee a = a$.

Esempio. L'insieme $\{0, 1\}$ costituisce la piú semplice algebra booleana $\mathbf{B}_1$ di due elementi con le operazioni $a \vee b = \max(a, b)$, $a \wedge b = \min(a, b)$, $a' = 1 - a$. In generale l'insieme $\mathbf{B}_n$ dei «codici binari» di lunghezza n, cioè delle successioni di n zeri/uni, forma un'algebra booleana di 2^n elementi. Per addizionare o moltiplicare due siffatte successioni occorre compiere la corrispondente operazione sui simboli binari di ciascun posto, e scrivere il risultato nello stesso posto. Per esempio, $(0110) \vee (1100) = (1110)$, $(010) \wedge (101) = (000)$.

Gli schemi elettronici che realizzano tali operazioni booleane erano tra le fondamentali unità di struttura dei calcolatori delle prime generazioni.

Nell'introduzione si è già parlato dei legami fra la numerazione binaria e la classificazione dicotomica. Se si interpreta il segno 1/0 nella i-esima posizione come indicazione di presenza/assenza dell'i-esimo contrassegno, allora ogni elemento dell'algebra booleana $\mathbf{B}_n$ descrive un certo insieme di oggetti classificabili: per esempio (0110) corrisponde agli oggetti che posseggono il secondo e il terzo contrassegno, ma non il primo e il quarto.

Esempio. Se S è un insieme qualsiasi, allora l'insieme di tutte le sue parti $B = \mathcal{P}(S)$ forma un'algebra booleana con dati strutturali $\{0 = \varnothing,\, 1 = S;\, \cap, \cup;$ complemento insiemistico in $S\}$.

Esempio. Se S è uno spazio topologico, allora il sistema dei suoi insiemi aperti regolari, cioè gli insiemi aperti con la condizione $a'' = a$, dove l'elemento a' è definito come l'insieme dei punti interni al complemento di a, è un'algebra booleana. Le operazioni $\vee$, $\wedge$ coincidono con $\cup$, $\cap$ rispettivamente.

Nel seguito è descritto il legame di questi esempi rispettivamente con la logica classica e con quella intuizionista.

A fondamento della dualità booleana sta la seguente osservazione: se negli enunciati degli assiomi 1-5 si scambiano di posto dovunque 0 con 1, e $\vee$ con $\wedge$, si ottiene un sistema di assiomi equivalente.

Questa osservazione conduce ad alcune conseguenze. La prima di queste è il principio di dualità nella metateoria delle algebre booleane, che permette senza sforzo di «raddoppiare» il numero delle identità booleane dimostrabili. Invero, sia data una qualsiasi identità booleana di validità generale, cioè che rimane vera comunque si sostituiscano le lettere che compaiono in essa con elementi arbitrari di una qualsiasi algebra booleana. Allora, invertendo in essa 0 e 1, e $\wedge$ e $\vee$, si ottiene ancora un'identità di validità generale.

Questo principio, che è un corollario della simmetria della teoria, è visibile già a livello di assiomi. Non ci sono molti esempi del genere; si può ancora citare la dualità proiettiva e la dualità delle iper-algebre (cfr. § 1.4).

Un'altra conseguenza è la possibilità di costruire per ogni algebra booleana $\mathbf{B}=\{B;\, 0, 1;\, \wedge, \vee;\, '\}$ l'algebra booleana ad essa *duale* $\mathbf{B}^*=\{B^*;\, 0^*, 1^*;\, \wedge^*, \vee^*;\, '^*\}$, dove $B^*=B$, $0^*=1$, $\wedge^*=\vee$, $\vee^*=\wedge$, $'^*='$. La simmetria degli assiomi mostra che $\mathbf{B}^*$ è un'algebra booleana; evidentemente $(\mathbf{B}^*)^*=\mathbf{B}$.

L'esistenza di un oggetto duale è un aspetto caratteristico di una serie di teorie. Tuttavia è raro che esso sia canonicamente isomorfo all'oggetto iniziale. Nella teoria delle algebre booleane accade proprio cosí, nel che si manifestano le proprietà di simmetria dell'operazione $'$.

Invero, l'applicazione $\mathbf{B} \to \mathbf{B}^*: a \mapsto a'$ è un isomorfismo dell'algebra $\mathbf{B}$ con l'algebra duale $\mathbf{B}^*$.

Se s'interpreta $'$ come un'operazione di complemento (per esempio nelle algebre $\mathcal{P}(S)$), allora diventa chiaro fondamento significativo del principio di dualità: il complemento insiemistico scambia di posto $0=\varnothing$ con $1=S$ e $\cup$ con $\cap$.

La teoria delle algebre booleane, date dagli assiomi 1-5 come oggetti di teoria degli insiemi, ha un ulteriore aspetto logico. Essa si rivela, passando al linguaggio duale, un «calcolo degli enunciati» (alla nuova e non formale dualità linguaggio/modello che appare qui si tornerà nel § 2).

Algebre booleane e logica.

Linguaggio degli enunciati.

In logica formale i segni $\wedge$, $\vee$, $'$ sono usati per denotare connettivi logici: le operazioni «e», «o» (non esclusivo: vel), «non», usate per congiungere alcuni enunciati in uno nuovo. Il linguaggio formale degli enunciati consiste di espressioni in un alfabeto finito, ottenute da espressioni di una sola lettera 0 (falso), 1 (vero), $p_1, \ldots, p_n$ (enunciati atomici) mediante applicazioni successive di queste operazioni; ad esempio $((p_1 \wedge p_2) \vee p_3)'$. Alcune espressioni risultano equivalenti dal punto di vista logico, per esempio $p_1 \wedge p_2$ e $p_2 \wedge p_1$. Proprio nella descrizione di una tale equivalenza consiste la descrizione della logica del linguaggio. Si considereranno brevemente due logiche: quella classica e quella intuizionista. La logica classica modella il meccanismo delle inferenze deduttive, utilizzate dalla matematica consueta. La logica intuizionista modella una relazione piú ristretta di equivalenza logica, elaborata come risultato dell'analisi degli aspetti positivi della critica intuizionista ai fondamenti della matematica, iniziata da Kronecker e da Brouwer. Dal punto di vista dell'argomento di questo

articolo, la differenza fondamentale tra la logica intuizionista e quella classica consiste nel rifiuto della simmetria della negazione: l'enunciato p'' è logicamente equivalente a p da un punto di vista classico, ma non dal punto di vista intuizionista. In altre parole, nella logica intuizionista non è vera la legge del terzo escluso.

Equivalenza logica nella logica classica. Siano P e Q due enunciati formati da enunciati atomici $p_1, ..., p_n$. Si consideri un'algebra booleana del tipo $\mathcal{P}(S)$ e si attribuiscano a $p_1, ..., p_n$ degli arbitrari valori di $\mathcal{P}(S)$. Interpretando poi $\wedge$, $\vee$ e $'$ come le operazioni $\cap$, $\cup$ e complemento in $\mathcal{P}(S)$, si possono assegnare a P e Q dei valori in $\mathcal{P}(S)$. Gli enunciati P e Q sono equivalenti dal punto di vista della logica classica se per qualunque S e per qualsiasi valore di $p_1, ..., p_n$ in S i valori di P e Q sono gli stessi.

Equivalenza logica nella logica intuizionista. Siano P e Q due enunciati formati da enunciati atomici $p_1, ..., p_n$ come sopra. Sia S uno spazio topologico. Si consideri l'insieme $\mathcal{I}(S)$ degli aperti di S, e si assegnino a $p_1, ..., p_n$ dei valori qualsiasi da $\mathcal{I}(S)$. Interpretando poi $\wedge$, $\vee$ e $'$ come le operazioni $\cap$, $\cup$ e «l'insieme dei punti interni del complemento» in $\mathcal{I}(S)$, si possono associare a P e Q dei valori in $\mathcal{I}(S)$. Gli enunciati P e Q sono equivalenti secondo la logica intuizionista se per ogni spazio topologico S e per valori arbitrari di $p_1, ..., p_n$ in $\mathcal{I}(S)$ i valori associati a P e Q sono gli stessi.

Da questa descrizione si può evincere il seguente risultato di Gödel: nell'ambito degli enunciati consistenti in doppie negazioni la logica intuizionista coincide con quella classica.

Dualità logica.

Se P è un enunciato (nel linguaggio degli enunciati), allora l'enunciato ad esso duale P^* si ottiene per sostituzione di $\wedge$ con $\vee$, di 0 con 1, di tutti gli enunciati atomici che entrano in P mediante le loro negazioni, e negando il risultato. (Nel linguaggio dei predicati, che ha come strumenti di espressione anche i quantificatori universale $\forall$ ed esistenziale $\exists$, occorre anche sostituire $\forall$ con $\exists$, e viceversa). Nella logica classica P e P^* sono logicamente equivalenti, nella logica intuizionista ciò non è vero.

Asimmetria della negazione.

La critica intuizionista mette in luce caratteristiche di asimmetria della negazione, o passaggio al complemento, significative, in una opportuna interpretazione, anche nella matematica classica.

Un esempio tipico di questa asimmetria si manifesta nella teoria dei sottoinsiemi computabili dell'insieme dei numeri naturali $\mathbf{N}$. Diremo che un sottoinsieme $E \subset \mathbf{N}$ è computabile se esiste un algoritmo che assegna gli elementi di E in un certo ordine. Si può dimostrare che esiste un insieme computabile E tale che il suo complemento non è computabile. In questo modo l'operazione di complemento porta fuori dai limiti dell'universo degli insiemi computabili, che da molti punti di vista è l'ambiente naturale per la matematica algoritmica.

Un modello analogo di asimmetria del passaggio al complemento si ottiene, nell'ambito della teoria classica degli insiemi, se si considera come universo la classe V di «tutti» gli insiemi di Neumann-Zermelo-Fraenkel. Il complemento di un insieme di questa classe è una classe, e non un insieme, cosicché $V^* \neq V$.

L'essenza di questa critica consiste nel fatto che anche nei nostri modelli di pensiero l'universo idealizzato nel quale lavoriamo è spesso incomparabilmente piú grande del dominio degli oggetti sottoposti realmente ad investigazione, e il complementare di questa regione risulta quasi ipso facto inaccessibile.

Il trascurare questa critica nel nome dei vantaggi di una dualità ideale sembra essere una delle importanti sorgenti della forza della matematica classica.

1.2. Dualità degli spazi lineari.

Questo paragrafo è dedicato ad una delle costruzioni piú importanti. Sia per il numero delle applicazioni e dei casi particolari che per la struttura generale, la dualità degli spazi lineari può essere considerata il prototipo delle dualità matematiche. Nel § 3 vi si ritornerà, considerando nella discussione anche gli aspetti topologici.

Linearità e spazi lineari. Sia k un campo (per esempio il campo reale $\mathbf{R}$ o il campo complesso $\mathbf{C}$). Si ricordi che uno spazio lineare (o vettoriale) L su k è un insieme di elementi (vettori) che possono essere addizionati tra loro e moltiplicati per elementi di k. Queste operazioni sono regolate da semplici assiomi.

Esempio. Sia X un insieme. Le funzioni su X a valori in k formano uno spazio vettoriale su k. Se $X = \{1, ..., n\}$, una qualsiasi funzione di questo tipo può essere rappresentata come una famiglia indicizzata di numeri, o vettore n-dimensionale $(a_1, ..., a_n)$, $a_i \in k$. Lo spazio vettoriale corrispondente si indica con k^n, ed è chiamato spazio coordinato n-dimensionale (su k). Lo spazio lineare L si dice a dimensione finita se esiste un isomorfismo di L con k^n, per un n opportuno, e a dimensione infinita in caso contrario. Questo numero n, se esiste, è definito univocamente e si dice dimensione di L. La scelta di un isomorfismo di L con k^n equivale all'introduzione di un sistema di coordinate in L. Si può sostituire questa scelta con la scelta di una base in L, cioè di una famiglia di vettori $(l_1, ..., l_n)$, $l_i \in L$ tale che ogni vettore di L sia esprimibile univocamente nella forma $\sum_{i=1}^{n} a_i l_i$, $a_i \in k$.

Probabilmente nelle scienze naturali l'aspetto piú importante dell'idea di linearità è il principio della linearità dei piccoli incrementi. Nella sua forma piú semplice esso postula che una funzione di argomento reale $f(x)$ goda della proprietà: $f(x_0+\varepsilon)-f_0(x) = \text{cost}\cdot\varepsilon +$ (infinitesimi di ordine superiore rispetto ad ε), per $\varepsilon \to 0$. Questo è il fondamento del calcolo differenziale e variazionale.

Il principio di sovrapposizione in meccanica quantistica, che introduce una struttura vettoriale complessa sullo spazio delle funzioni d'onda di ogni sistema fisico, ha elevato ancora di piú lo status ontologico della linearità. Nel suo linguaggio, in particolare, si dà una descrizione precisa del dualismo onda/corpusco-

lo (cfr. anche § 3). Infine, nei modelli fondamentali della fisica (meccanica galileiana, teoria della relatività ristretta), lo spazio-tempo è lineare su **R**, e nei modelli della relatività generale esso è «localmente in piccolo» lineare su **R**, cioè è una varietà differenziabile. Il principio di linearità dei piccoli incrementi è descrivibile nel linguaggio della geometria differenziale.

Esempio: spazio tangente in un punto. Sia X una varietà differenziabile. Lo spazio T_xX tangente ad essa nel punto $x \in X$ è quello spazio vettoriale sui numeri reali che può essere definito nel modo seguente. Sia $f:(0,1) \to X$ un'applicazione liscia, cioè una «curva», passante per il punto x. Per ogni funzione liscia F su X si può definire la derivata di F in x lungo f come il limite

$$\lim_{\Delta t \to 0} \frac{F \circ f(t_0+\Delta t) - F \circ f(t_0)}{\Delta t}$$

se $f(t_0)=x$. Si chiama vettore tangente $\xi \in T_xX$ la classe di tutte le curve f aventi gli stessi valori delle derivate per tutte le F. Intuitivamente un vettore tangente definisce ed è definito dalla classe di curve lisce passanti per il punto x con la stessa direzione e con la stessa velocità istantanea.

Il vettore tangente $\xi \in T_xX$ definisce una derivazione dell'anello delle funzioni lisce su X a valori in $\mathbf{R}$: $\xi(aF+bG)=a\xi F+b\xi G$, $\xi(FG)=G(x)\xi F+F(x)\xi G$, dove $a,b \in \mathbf{R}$ e ξF è la derivata di F lungo ξ.

Si ricordi che, se L e M sono due spazi vettoriali su k, l'applicazione $f: L \to M$ è detta (k)-lineare se $f(al_1+bl_2)=af(l_1)+bf(l_2)$ per ogni $a,b \in k$; $l_1, l_2 \in L$.

Esempio. Se X, Y sono varietà differenziabili, $f: X \to Y$ un'applicazione liscia, $x \in X$, $y=f(x)$, allora è definita un'applicazione lineare $(df)_x: T_xX \to T_yY$ degli spazi tangenti. È questo il modello standard del principio di «linearità dei piccoli incrementi»; $(df)_x$ è detto differenziale dell'applicazione f nel punto $x \in X$.

Spazio duale. Sia L uno spazio vettoriale su k. Si dice funzionale lineare su L un'applicazione lineare $f: L \to k$. L'insieme di tutti i funzionali lineari $\mathfrak{L}(L,k)$ su L si indica con L^* e si chiama spazio duale o coniugato di L.

Esempio. Lo spazio $(T_xX)^* = T_x{}^*X$, duale dello spazio tangente nel punto x, si chiama cotangente. Ogni funzione liscia f su X definisce un elemento dello spazio cotangente $(df)_x \in T_x{}^*X$, per il quale $(df)_x(\xi)=\xi f$ se $\xi \in T_xX$. (Si ricordi che ξf è la derivata di f nella direzione ξ).

Esempio. Sia $C[0,1]$ lo spazio vettoriale delle funzioni continue su $[0,1]$. L'integrale di Riemann $f \mapsto \int_0^1 f(x)\,dx$ è un funzionale lineare su $C[0,1]$, e perciò un elemento di $C[0,1]^*$. Fissato $g \in C[0,1]$, si può considerare anche il funzionale $f \mapsto \int_0^1 f(x)g(x)\,dx$. Diverse varianti di questa costruzione conducono, ad esempio, alla teoria della misura nel linguaggio della dualità degli spazi funzionali ($g(x)dx$ può essere sostituita da una misura piú generale $d\mu(x)$), ed anche alla teoria delle funzioni generalizzate (distribuzioni) (cfr. § 3).

Principî di dualità. Fissato un vettore $l \in L$, si consideri l'applicazione $L^* \to k$ che all'elemento $f \in L^*$ fa corrispondere $f(l) \in k$. L'applicazione risultante è lineare in f e perciò è un elemento dello spazio due volte coniugato $L^{**}=(L^*)^*$. A sua volta quest'elemento dipende linearmente da f; ciò definisce l'applicazione lineare canonica $\beta: L \to L^{**}$: $l \mapsto \{f \mapsto f(l)\}$.

Il primo principio di dualità lineare è il seguente: L'applicazione canonica $\beta: L \to L^{**}$ è un'applicazione lineare iniettiva. Essa è un isomorfismo di spazi vettoriali se L è a dimensione finita.

La linearità di β è evidente dalla definizione. L'iniettività sta ad indicare che per ogni $l \in L$ esiste $f \in L^*$ con $f(l) \neq 0$. Ciò si verifica facilmente per induzione transfinita, prolungando f su sottospazi crescenti di L contenenti l. Che β sia un isomorfismo nel caso a dimensione finita si verifica facilmente servendosi di una base in L, cioè di un sistema di elementi $\{l_1, \ldots, l_n\} \in L$ tale che ogni $l \in L$ si esprime univocamente nella forma $\sum_{i=1}^{n} a_i l_i$, $a_i \in k$. Allora i funzionali $l_i{}^* \in L^*$, per i quali $l_i{}^*(l_j)=0$ per $i \neq j$ e 1 per $i=j$ formano una base di L^* e, inoltre, $\beta(l_i)=(l_i{}^*)^*$, dove $(l_i{}^*)^*$ è la base di L^{**} costruita in modo analogo.

Nel caso a dimensione infinita il «biduale» L^{**} costituisce una estensione non banale di L (cfr. § 3). Un uso analogo della doppia «dualizzazione» per la costruzione di oggetti generalizzati s'incontra spesso nella sua variante categoriale (prefasci su categorie: cfr. § 2).

Siano L, M due spazi vettoriali, $\varphi: L \to M$ un'applicazione lineare. A partire da questa si definisce l'applicazione aggiunta $\varphi^*: M^* \to L^*$, che al funzionale $f: M \to k$ fa corrispondere l'applicazione composta $f \circ \varphi: L \to k$. Si vede facilmente che φ^* è lineare se lo è φ, e che $(\psi \circ \varphi)^* = \varphi^* \circ \psi^*$. Applicando ancora una volta questa operazione si ottiene l'applicazione lineare $\varphi^{**}: L^{**} \to M^{**}$.

Il secondo principio di dualità lineare è il seguente: Nell'identificazione canonica di L^{**} con L e di M^{**} con M (L ed M a dimensione finita), l'applicazione φ^{**} si identifica con φ.

Se negli spazi L, M si fissano le basi (l_i), (m_j), allora le applicazioni φ sono in corrispondenza biunivoca con le matrici di dimensione $(\dim L, \dim M)$ su k: alla matrice (a_{ij}) corrisponde l'applicazione $l_i \mapsto \sum_j a_{ij} m_j$. Allora nelle basi (m_j^*), (l_i^*) all'applicazione φ^* corrisponde la matrice (a_{ij}^t), dove $a_{ij}^t = a_{ji}$ («riflessione rispetto alla diagonale principale»), e l'applicazione φ^{**} corrisponde alla duplice riflessione di questa matrice, il che rende ovvia l'uguaglianza $\varphi^{**}=\varphi$.

Siano $L \xrightarrow{\varphi} M \xrightarrow{\psi} N$ applicazioni lineari tra spazi vettoriali. Si dice che esse formano un complesso se $\psi \circ \varphi$ è l'applicazione nulla, cioè se $\operatorname{Im}\varphi$ (immagine di φ) $\subseteq \operatorname{Ker}\psi$ (nucleo di ψ). Si dice che questo complesso è esatto in M se $\operatorname{Im}\varphi = \operatorname{Ker}\psi$. Una terminologia analoga si usa per le coppie di applicazioni consecutive in successioni di lunghezza arbitraria $\ldots \to L_{i-1} \to L_i \to L_{i+1} \to \ldots$ Per esempio, il fatto che la successione $0 \to L \xrightarrow{\varphi} M$ sia esatta significa che φ è una iniezione di L in M, e che lo sia la successione $M \xrightarrow{\psi} N \to 0$ indica che ψ è una surie-

zione di M su N. (Qui o è lo spazio vettoriale consistente del solo vettore nullo). Ogni applicazione lineare $L \xrightarrow{\varphi} M$ è inclusa in una successione esatta $0 \to \mathrm{Ker}\,\varphi \to L \xrightarrow{\varphi} M \to \mathrm{Coker}\,\varphi \to 0$, dove, per definizione, $\mathrm{Coker}\,\varphi = M/\mathrm{Im}\,\varphi$ è lo spazio quoziente di M rispetto all'immagine di φ.

L'ultimo, importante principio di dualità lineare che si ritiene di dover indicare in questa sede può essere formulato nel modo seguente: Se la successione $L \xrightarrow{\varphi} M \xrightarrow{\psi} N$ è esatta in M, allora la successione $N^* \xrightarrow{\psi^*} M^* \xrightarrow{\varphi^*} L^*$ è esatta in M^*.

Nel § 2 si darà una sintetica formulazione categoriale che unifica tutti questi principî.

Sintesi di L e L^*. A differenza di quanto avviene nella dualità booleana, non esiste un isomorfismo canonico fra gli spazi vettoriali L e L^*. La «sintesi» di L e dello spazio ad esso duale L^* deve effettuarsi mediante un'ulteriore struttura esterna.

Tra le piú importanti di tali strutture figurano l'assegnazione di un prodotto scalare in L e la scelta di una particolare base in L.

Esempio: scelta di una base. Sia L uno spazio a dimensione finita, e sia fissata in esso una base $(l_1, \ldots, l_n)$. Allora in L^* è determinata una base (l_i^*), dove $l_i^*(l_j) = \delta_{ij}$, cosa che permette di definire un isomorfismo $L \simeq L^* : l_i \mapsto l_i^*$, $i = 1, \ldots, n$. Un caso tipico in cui si ha una base assegnata è lo spazio delle funzioni su un insieme finito, per esempio l'anello di gruppo di un gruppo finito (cfr. oltre).

Esempio: prodotto scalare. Sia E uno spazio euclideo a dimensione finita. Ciò significa che, comunque dati due vettori $x, y \in E$, è definito il loro prodotto scalare $(x,y) \in \mathbf{R}$, lineare in x e y; (x,x) è il quadrato della lunghezza del vettore x. L'applicazione $E \to E^*$, che al vettore x fa corrispondere il funzionale $y \mapsto (x,y)$ è un isomorfismo canonico di E con E^*. La struttura esterna supplementare che permette di costruire tale isomorfismo è il prodotto scalare.

L'assiomatizzazione delle proprietà caratteristiche di questo esempio conduce all'importante concetto di applicazione e forma bilineare su uno spazio vettoriale. L'isomorfismo di uno spazio con il suo duale in tale contesto si realizza abitualmente mediante forme di una fra due classi: ortogonale o simplettica.

Applicazioni bilineari.

Siano L, M, N spazi vettoriali. L'applicazione $f: L \times M \to N$ si dice bilineare se è lineare in ciascuno dei due argomenti $l \in L$, $m \in M$, quando l'altro è fissato. Un esempio è l'applicazione bilineare canonica $L^* \times L \to k : (l^*, l) \mapsto l^*(l)$. Nucleo sinistro (risp. destro) di un'applicazione bilineare $f: L \times M \to N$ si dice l'insieme di quei vettori $l \in L$ (risp. $m \in M$) tali che $f(l,m) = 0$ per ogni $m \in M$ (risp. per ogni $l \in L$). Un'applicazione bilineare f si dice (algebricamente) non-degenere (risp. non-degenere a sinistra, a destra), se entrambi i suoi nuclei (risp. sinistro, destro) sono uguali a zero. L'applicazione bilineare canonica $L^* \times L \to k$ è non-degenere.

L'applicazione bilineare $f: L \times M \to k$ definisce un'applicazione lineare $L \to M^* : l \mapsto \{m \mapsto f(l,m)\}$ e $M \to L^* : m \mapsto \{l \mapsto f(l,m)\}$. Se f è non-degenere a sinistra, $L \to M^*$ è una immersione; se f è non-degenere a destra, $M \to L^*$ è una immersione. Da ciò segue che se f è non-degenere, le applicazioni $L \to M^*$ e $M \to L^*$ sono isomorfismi. Essi sono mutuamente coniugati nel senso definito sopra, cosí che l'assegnare un'applicazione bilineare non-degenere $f: L \times M \to k$ è equivalente ad assegnare una identificazione $L \simeq M^*$ oppure $M \simeq L^*$.

Si consideri ora il caso in cui $M = L$. Secondo quanto detto sopra, l'assegnazione di un isomorfismo $L \simeq L^*$ è equivalente all'assegnazione di un'applicazione bilineare non degenere $L \times L \to k$ o forma bilineare sullo spazio L. Nello spazio euclideo si ha una forma bilineare canonica: il prodotto scalare di vettori; perciò lo spazio euclideo è identificato canonicamente con il suo duale, cosí come avveniva nell'esempio portato sopra.

Forme ortogonale e simplettica.

La forma bilineare $f: L \times L \to k$ può essere analizzata rispetto a un altro tipo di simmetria: quella relativa alla permutazione dei suoi argomenti. Se $f(l_1, l_2) = f(l_2, l_1)$ per tutti gli $l_1, l_2 \in L$, la forma si dice simmetrica; se invece $f(l_1, l_2) = -f(l_2, l_1)$, si dice antisimmetrica e, nel caso non-degenere, simplettica. Una qualsiasi forma f si decompone nella somma di una componente simmetrica e di una antisimmetrica $f_+ + f_-$:

$$f_+(l_1, l_2) = \frac{1}{2}(f(l_1, l_2) + f(l_2, l_1))$$

$$f_-(l_1, l_2) = \frac{1}{2}(f(l_1, l_2) - f(l_2, l_1)).$$

La teoria delle coppie formate da uno spazio L e da una forma non-degenere f assegnata su di esso viene detta geometria ortogonale (astratta) se f è simmetrica, geometria simplettica se f è antisimmetrica. Sebbene per noi al centro dell'attenzione si trovi l'identificazione di L con L^*, definita da f, l'intuizione geometrica originaria si basa sulla rappresentazione di $f(l,l)$ come «quadrato della lunghezza» del vettore l, in analogia con il caso euclideo $k = \mathbf{R}$, $L = \mathbf{R}^3$, $f((x_1, x_2, x_3), (x_1, x_2, x_3)) = \sum_{i=1}^{3} x_i^2$. Ciò conduce, per esempio, allo studio del gruppo dei movimenti dello spazio L, gli automorfismi lineari che conservano la «metrica» $\|l\|^2 = f(l,l)$, ecc.

Sebbene storicamente la geometria ortogonale si sia trovata a lungo al centro dell'attenzione, a poco a poco le strutture simplettiche si sono guadagnate un ruolo di parità e si riconosce il loro ruolo fondamentale proprio nelle situazioni in cui oggetti duali sono congiunti in uno unico.

Esempio: geometria simplettica su $L^* \times L$. Sia L uno spazio vettoriale, L^* il suo duale. La somma diretta $L^* \times L$ consiste delle coppie (f, l), $f \in L^*$, $l \in L$. Su $L^* \times L$ si ha una forma antisimmetrica canonica non-degenere $\langle , \rangle$: $\langle (f_1, l_1), (f_2, l_2) \rangle = f_1(l_2) - f_2(l_1)$. La restrizione di questa forma ai sotto-

spazi $L, L^* \subset L^* \times L$ è identicamente nulla. Inoltre L ed L^* sono sottospazi massimali con questa proprietà. Si può dimostrare che ogni sottospazio a dimensione finita M, su cui sia data una forma simplettica non-degenere, è isomorfo ad uno spazio del tipo $L^* \times L$. La scelta di un tale isomorfismo equivale alla scelta di due sottospazi lagrangiani complementari in M: i sottospazi di dimensione massimale sui quali la restrizione della forma simplettica sia identicamente nulla.

Esempio: algebra di Grassmann. Siano k un campo e $x_1, \ldots, x_n$ n variabili indipendenti su k. Si consideri un anello (non commutativo) Λ su k, generato da $x_1, \ldots, x_n$, con le relazioni anticommutative $x_i x_j = -x_j x_i$ per ogni $i, j = 1, \ldots, n$ (in particolare, $x_i^2 = 0$). Se tutte le rimanenti relazioni sono conseguenza di queste e degli assiomi di k-algebra, allora Λ si chiama algebra di Grassmann con un sistema di generatori fissato $x_1, \ldots, x_n$. Sia $\Lambda^i \subset \Lambda$ il sottospazio formato dai polinomi omogenei di grado i, $\Lambda^0 = k$. Si può verificare che la dimensione di Λ^i è uguale a

$$\binom{n}{i} = \frac{n(n-1)\ldots(n-i+1)}{i!},$$

in particolare $\Lambda^n = k x_1 \ldots x_n$, e $\Lambda^j = 0$ per $j > n$. La moltiplicazione $\Lambda^i \times \Lambda^{n-i} \to \Lambda^n$ definisce una forma bilineare non-degenere φ_i: $ab = \varphi_i(a, b) x_1 \ldots x_n$, ove $a \in \Lambda^i$, $b \in \Lambda^{n-i}$. Perciò Λ^i s'identifica con $(\Lambda^{n-i})^*$. Sulla parte «dispari» dell'algebra Λ^{2i+1} la forma φ_{2i+1} è antisimmetrica se n è pari, sulla parte pari Λ^{2i} è simmetrica.

L'algebra di Grassmann senza generatori fissati può essere definita come l'algebra generata dalle applicazioni multilineari antisimmetriche

$$f : \underbrace{L \times \ldots \times L}_{i} \to k,$$

dove L è uno spazio a dimensione finita. Essa si indica con ΛL^*.

Nei modelli fisici i polinomi commutativi e anticommutativi corrispondono a particelle elementari con statistiche diverse: bosoni e fermioni. La teoria unificata di entrambi i tipi di particelle richiede l'esame di algebre una parte dei cui generatori è commutativa, mentre l'altra è anticommutativa, e dell'interessante nuovo apparato della «supersimmetria».

Nella geometria e nella topologia differenziali hanno un ruolo importante le varietà differenziabili X sui cui spazi tangenti $T_x X$, $x \in X$, siano date strutture del tipo del prodotto scalare, che variano in modo liscio con x.

La regolarità rispetto ad x qui e nel seguito va intesa nel modo seguente: l'unione di tutti gli spazi tangenti $TX = \bigcup_{x \in X} T_x X$ può essere dotata di una struttura di varietà differenziabile in modo tale che la proiezione $\pi : TX \to X$ (che ad un vettore tangente nel punto x associa il punto x) sia un'applicazione liscia. Un'osservazione analoga si può fare per $T^*X = \bigcup_{x \in X} T_x^* X$, $TX \times_X TX = \bigcup_{x \in X} (T_x X \times \times T_x X)$, ecc. TX prende il nome di fibrato tangente ad X, e T^*X di fibrato cotangente.

Una sezione liscia dell'applicazione $\pi : TX \to X$ si chiama allora campo vettoriale, una sezione liscia di $\pi^* : T^*X \to X$, forma differenziale (di grado 1) su X. Gli spazi $D(X)$ dei campi vettoriali, e $\Omega^1(X)$ delle forme differenziali sono lineari e ad infinite dimensioni; su di essi è definita un'applicazione bilineare canonica $D(X) \times \Omega^1(X) \to C^\infty(X)$ nello spazio delle funzioni lisce, cioè dei campi scalari su X.

Se sulle fibre di TX è data una forma ortogonale (risp. simplettica) non-degenere che dipende in modo liscio da x, allora X viene detta varietà pseudoriemanniana (risp. simplettica).

Una varietà pseudoriemanniana si dice riemanniana se tale forma è definita positiva. Su una varietà pseudoriemanniana, in un sistema locale di coordinate $(x^1, \ldots, x^n)$, il «quadrato della lunghezza» di un piccolo vettore tangente con origine in $(x^1, \ldots, x^n)$ e fine in $(x^1 + dx^1, \ldots, x^n + dx^n)$, per tradizione si indica con $\sum g_{ij} dx^i dx^j$.

Esempio: strutture riemanniana e simplettica in meccanica. Sia X lo spazio delle configurazioni di un sistema meccanico, TX il suo fibrato tangente. Se $(q_1, \ldots, q_n)$ sono coordinate locali su X, le coordinate locali su TX si esprimono in genere nella forma $(q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n)$: nel vettore tangente $\xi \in T_x X$ il valore di q_i è uguale a $q_i(x)$, e il valore di $\dot{q}_i$ è uguale a ξq_i. L'energia cinetica del sistema spesso ha la forma $\sum a_{ij}(q_1, \ldots, q_n) \dot{q}_i \dot{q}_j$, cioè dipende quadraticamente dalla velocità. Se questa forma è positiva e non-degenere, essa definisce una metrica riemanniana $\sum a_{ij} dq_i dq_j$ sullo spazio delle configurazioni X.

Le equazioni di evoluzione del sistema in forma hamiltoniana si scrivono per una traiettoria nello spazio delle fasi di questo sistema, che è definito come il *fibrato cotangente* T^*X. Queste equazioni si scrivono in termini della struttura simplettica canonica su T^*X. Il meccanismo della sua costruzione è analogo a quello del caso $L^* \times L$: lo spazio lineare $M = T_y(T^*X)$ contiene $T_x^* X$, dove x è l'immagine di y su X, in qualità di sottospazio, e lo spazio quoziente rispetto ad esso è isomorfo a $T_x X$. L'assenza di una decomposizione diretta è compensata dalla condizione che la forma simplettica sia liscia e chiusa.

Questa forma simplettica ha un ruolo fondamentale nella formulazione delle equazioni di evoluzione della meccanica hamiltoniana.

Esempio: dualità di Poincaré. In topologia si ha una serie di costruzioni che permettono di associare a uno spazio topologico X e a un campo k il suo anello di coomologia a coefficienti in k: $H^*(X, k)$. Per esempio, se X è una varietà differenziabile, $k = \mathbf{R}$, si può considerare il cosiddetto complesso di De Rham su X. Esso è una successione di spazi lineari $\Omega^i X$ (le sezioni lisce del fascio $\Lambda^i T^*X$) e di applicazioni $d_i : \Omega^i X \to \Omega^{i+1} X$, che qui non si definiranno. Allora per definizione $H^*(X, \mathbf{R}) = \bigoplus_{i=0}^{\infty} H^i(X, \mathbf{R})$ e $H^i(X, \mathbf{R}) = \operatorname{Ker} d_i / \operatorname{Im} d_{i-1}$. La moltiplicazione in $H^*(X, \mathbf{R})$ è indotta dalla moltiplicazione grassmanniana, o esterna, in $\Lambda^i T^*(X)$. Se $n = \dim X$, e X è connesso ed orientato, allora $H^n(X) \simeq \mathbf{R}$, e la scelta di questo isomorfismo definisce una forma bilineare

non-degenere $H^i(X)\times H^{n-i}(X)\to\mathbf{R}$, che fornisce anche uno degli aspetti della dualità di Poincaré: $H^i(X)\simeq(H^{n-i}(X))^*$. Non si può qui accennare all'altro aspetto, la dualità fra coomologia ed omologia, ed anche alle sue molteplici varianti e generalizzazioni.

Algebra tensoriale e dualità. A partire da due spazi vettoriali L ed M, se ne può costruire un terzo, il loro prodotto tensoriale $L\otimes M$. Questa costruzione ha un ruolo sempre crescente nella matematica contemporanea. Oltre alle applicazioni classiche nella teoria delle rappresentazioni e nella geometria differenziale, i prodotti tensoriali sono stati molto usati nell'algebra commutativa: un prodotto tensoriale di anelli si trasforma nel prodotto diretto dei loro spettri nella categoria duale (cfr. § 2). Il formalismo dei prodotti tensoriali viene usato in modo essenziale nella descrizione quantistica dell'unione di piú sistemi: la funzione d'onda di un sistema unione di piú sistemi appartiene allo spazio prodotto tensoriale degli spazi delle funzioni d'onda relative alle sue parti.

In questo paragrafo si descriverà il linguaggio tensoriale della geometria differenziale classica e si illustrerà il ruolo della dualità lineare in operazioni sui tensori quali l'abbassamento o innalzamento degli indici e la contrazione.

Definizione di prodotto tensoriale.

Siano L, M due spazi vettoriali sul campo k. Si consideri lo spazio N, con base $\{(l,m)\}$, dove $l\in L$, $m\in M$ sono tutte le possibili coppie di vettori. Nello spazio N si consideri il sottospazio N_0 generato dagli elementi $(al,m)-a(l,m)$, $(l_1+l_2,m)-(l_1,m)-(l_2,m)$, dove $a\in k$; $l,l_i\in L$, $m\in M$; e, analogamente, dagli elementi $(l,bm)-b(l,m)$, $(l,m_1+m_2)-(l,m_1)-(l,m_2)$. Posto $L\otimes M=N/N_0$ si definisce l'applicazione canonica $\gamma: L\times M\to L\otimes M$, $\gamma(l,m)=$ classe (l,m) mod $N_0=l\otimes m$.

L'applicazione γ è bilineare e gode delle seguenti proprietà generali: per ogni applicazione bilineare $\varphi: L\times M\to N$ esiste un'unica applicazione lineare $\psi: L\otimes M\to N$ tale che $\varphi=\psi\circ\gamma$.

Proprietà elementari.

Se (l_i), (m_j) sono basi di L, M, allora il sistema $(l_i\otimes m_j)$ forma una base di $L\otimes M$.

Sono definite le identificazioni canoniche di $L\otimes M$ con $M\otimes L$: $l\otimes m\mapsto m\otimes l$ e di $(L\otimes M)\otimes N$ con $L\otimes(M\otimes N)$: $l\otimes(m\otimes n)\mapsto(l\otimes m)\otimes n$, che esprimono la commutatività e l'associatività del prodotto tensoriale.

Ha luogo anche l'importante isomorfismo canonico di spazi vettoriali (nel caso a dimensione finita), funtoriale rispetto ad L, M (cfr. § 2): $\mathrm{Hom}(L,M)=L^*\otimes M$. All'elemento $l^*\otimes m\in L^*\otimes M$ esso associa l'applicazione lineare $L\to M: l\mapsto l^*(l)m$. Si scriverà $L^{\otimes p}$ in luogo di $L\otimes\ldots\otimes L$ (p volte).

Tensori.

Un elemento dello spazio $L^{\otimes p}\otimes L^{*\otimes q}$ si chiama tensore sullo spazio L, p volte covariante e q volte controvariante, di rango $p+q$. L'intera somma

$\bigoplus_{p,q\geq 0} L^{\otimes p}\otimes L^{*\otimes q}$ si dice algebra tensoriale su L (la moltiplicazione è quella tensoriale, con l'evidente commutatività dei fattori).

L'assegnare una forma non-degenere $\omega: L\times L\to k$ definisce sull'algebra tensoriale su L tutta una serie di strutture, che qui ora si elencherà con la nomenclatura classica. Sarà evidente che il ruolo di ω consiste proprio nel fissare un isomorfismo tra L e L^*.

Tensore metrico.

È la stessa forma ω, considerata come tensore di $(L\otimes L)^*\simeq L^*\otimes L^*$. Questa denominazione in genere si usa quando ω è simmetrica (su $\mathbf{R}$).

Innalzamento e abbassamento degli indici.

Dal momento che ω definisce un isomorfismo tra L e L^*, allora, sostituendo nel prodotto tensoriale $L\otimes L\otimes\ldots\otimes L^*\otimes L^*\otimes\ldots$ alcuni dei fattori L con L^* e viceversa, si ottiene una serie infinita di isomorfismi fra spazi di tensori, che non cambiano il rango totale. Tutti questi isomorfismi vengono detti innalzamenti e abbassamenti di indici, in conformità con la tradizionale notazione in termini di basi. Invero, se $\{l_i\}$ è una base di L, con $\{l^i\}$ si indica abitualmente la sua immagine in L^* rispetto all'isomorfismo scelto, e allora le operazioni d'innalzamento ed abbassamento degli indici trasformano, ad esempio, $l_1\otimes l_2$ in $l_1\otimes l^2$, $l^1\otimes l_2$ in $l^1\otimes l^2$, ecc. In tal modo ogni tensore, grazie alla forma ω, ottiene un'esistenza in diverse «ipostasi» che differiscono per i gradi di co- e controvarianza. In particolare il tensore metrico ω può essere considerato non solo come elemento di $L^*\otimes L^*$, ma anche di $L\otimes L^*$ o di $L\otimes L$.

Contrazione.

La forma bilineare simmetrica $L^*\times L\to k$ definisce un'applicazione lineare $L^*\otimes L\to k$. Scegliendo nel prodotto tensoriale $L\otimes L\otimes\ldots\otimes L^*\ldots$ due termini L e L^* (non necessariamente consecutivi), e applicando ad essi quest'applicazione lineare, si ottiene un'applicazione nei tensori di rango inferiore, detta contrazione rispetto alla corrispondente coppia di indici, uno in alto e l'altro in basso. Per esempio, la contrazione porta $l_i\otimes l^i$ in 1, e $l_i\otimes l^j$ (per $i\neq j$) in 0.

Esempio fondamentale.

Sia X una varietà differenziabile, TX il suo fibrato tangente. Nella teoria classica si chiamano tensori di rango (p,q) su X le sezioni del fibrato $T^{p,q}(X)=\bigcup_{x\in X}(T_xX)^{\otimes p}\otimes(T_x^*X)^{\otimes q}$, cioè i campi di tensori su T_xX che variano in modo liscio con x. L'assegnazione su X di una struttura (pseudo)riemanniana definisce un'identificazione fibra per fibra di T_xX con T_x^*X, e le corrispondenti operazioni di contrazione e d'innalzamento ed abbassamento di indici.

Se $(x^1,\ldots,x^n)$ è un sistema di coordinate locali su X, allora i campi vettoriali

$$\left(\frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_n}\right)$$

formano una base locale in T_xX, ed i differenziali $(dx_1, \ldots, dx_n)$ formano una base locale in T_x^*X. La notazione classica dei tensori è la notazione in termini delle loro coordinate in tali basi. Ad esempio, se f è una funzione su X, allora df è una sezione tensoriale in T^*X; nella notazione classica, $(\partial f/\partial x^\alpha)\, dx^\alpha$.

Si osservi in conclusione che si sono introdotti i concetti fondamentali dell'algebra tensoriale, ma non si è affrontata l'analisi tensoriale (connessione, curvatura, integrazione di forme, ecc.), in cui non è sufficiente considerare i tensori in ogni punto x separatamente, ma sono essenziali le proprietà delle loro variazioni con x.

Dualità e convessità.

Dualità di Legendre-Young.

Sia L uno spazio vettoriale a dimensione finita su k, e L^* il suo duale. Si scriverà (q,p) invece di $q(p)$ per $q \in L^*$, $p \in L$.

Le funzioni $f: L \to k$ e $g: L^* \to k$ si dicono duali secondo Young se per ogni punto $q_0 \in L^*$ (risp. $p_0 \in L$) esiste un unico punto $p_0 \in L$ (risp. $q_0 \in L^*$) tale che $f(p_0)+g(q_0)=(q_0,p_0)$. In una versione piú ristretta della definizione, questa condizione può essere soddisfatta solo quando (q_0,p_0) varia in un sottoinsieme di $L^* \times L$.

Nelle applicazioni il caso piú importante è quello in cui $k=\mathbf{R}$.

Sotto alcune condizioni su f si può dimostrare che la funzione ad essa duale g esiste, è unica, e soddisfa le stesse condizioni. Un esempio di tali condizioni è che f sia una funzione liscia e la regione $\{(x,y) \mid y \geq f(x)\} \subset L \times \mathbf{R}$ sia convessa, cioè contenga, insieme con due qualsiasi punti, tutto il segmento che unisce tali punti.

In tale caso f e g si chiamano trasformate di Legendre l'una dell'altra. Tali sono le coppie q^α/α, p^β/β per $\alpha, \beta > 1$, $(1/\alpha)+(1/\beta)=1$ (nella regione $q>0$, $p>0$).

L'assegnare una coppia di funzioni duali secondo Young f, g definisce un'applicazione biiettiva molto importante, ma in generale non-lineare:

$$L \to L^* : q_0 \mapsto p_0 \qquad \text{se } f(p_0)+g(q_0)=(q_0,p_0).$$

Esempio. Sia X lo spazio delle configurazioni di un sistema meccanico. La legge di evoluzione di questo sistema può essere data sia mediante la sua lagrangiana $\mathcal{L}$, sia mediante l'hamiltoniana $\mathcal{H}$. La prima è una funzione $\mathcal{L}: TX \to \mathbf{R}$, la seconda una funzione $\mathcal{H}: T^*X \to \mathbf{R}$. Per ogni $x \in X$ le restrizioni di $\mathcal{L}$ a T_xX e di $\mathcal{H}$ a T_x^*X rispettivamente costituiscono le trasformate di Legendre l'una dell'altra. La traiettoria del sistema in TX si trasforma nella sua traiettoria in T^*X, in seguito alla trasformazione sulle fibre $T_xX \to T_x^*X$ descritta sopra. Nei casi piú semplici la lagrangiana di un sistema è la differenza fra le sue energie cinetica e potenziale, e la hamiltoniana la loro somma.

Dualità delle figure convesse.

Sia di nuovo L uno spazio lineare a dimensione finita su $\mathbf{R}$, $E \subset L$ un insieme convesso chiuso contenente o come punto interno. L'insieme $E^* = \{p \in L^* \mid$

$\forall q \in E,\ (q,p) \leq 1\} \subset L^*$ si dice il duale di E. Se L è euclideo, in particolare è identificato canonicamente a L^*, allora E^*, come sottoinsieme di L, si dice polare di E.

Nel § 3 si vedrà che gli insiemi convessi sono usati per definire una topologia negli spazi ad infinite dimensioni, ed i loro duali definiscono la topologia corrispondente nello spazio duale.

Semilinearità e particolarità del caso complesso. Nel campo dei numeri complessi $\mathbf{C}$ si ha un unico automorfismo continuo non-banale: la coniugazione complessa. Nell'algebra lineare su $\mathbf{C}$ risulta assolutamente essenziale seguire l'azione della coniugazione complessa su tutti gli oggetti che s'incontrano. Ecco alcune definizioni.

Siano L, M due spazi vettoriali su $\mathbf{C}$. Un'applicazione $f: L \to M$ si dice semilineare se $f(l_1+l_2)=f(l_1)+f(l_2)$ e $f(al)=\bar{a}f(l)$ per qualsiasi $a \in \mathbf{C}$; $l, l_1, l_2 \in L$. Una terminologia analoga si adopera per le applicazioni dipendenti da piú argomenti vettoriali. Le forme che dipendono da due argomenti, linearmente da uno di essi e semilinearmente dall'altro, si chiamano sesquilineari.

La forma sesquilineare non-degenere $(x,y)=\sum_{i=1}^{n} x_i\bar{y}_i$ su $\mathbf{C}^n$ possiede le proprietà essenziali della metrica euclidea su $\mathbf{R}^n$, quali la realtà e la positività del quadrato della lunghezza di ogni vettore. Limitandosi alle forme bilineari sarebbe stato impossibile ottenere ciò.

Sia L uno spazio vettoriale su $\mathbf{C}$, e $L^{1,0}=L^*$, $L^{0,1}=$ spazio delle applicazioni semilineari di L in $\mathbf{C}$. Lo spazio $\bigoplus_{n=0}^{\infty} (L^{1,0} \times L^{0,1})^{\otimes n}$ appartiene all'algebra tensoriale ampliata di L: esso consiste dei tensori che si trasformano semilinearmente rispetto a certi indici, con un cambio del sistema di coordinate.

Applicando questa costruzione alle forme differenziali su una varietà complessa, si arriva all'importante concetto di forma di tipo (p,q) (per esempio, $dz_1 \ldots dz_p$, $d\bar{z}_1 \ldots d\bar{z}_q$) e di decomposizione di Hodge. Qui non c'è spazio per attardarsi su queste piú precisamente.

1.3. Gruppi abeliani finiti e loro caratteri.

Gruppo dei caratteri del gruppo G.

Sia G un gruppo, $\mathbf{C}^\cdot$ il gruppo moltiplicativo dei numeri complessi. Si dice carattere (unidimensionale, complesso, unitario) del gruppo G un omomorfismo $\chi: G \to \mathbf{C}^\cdot$ tale che $|\chi(g)|=1$ per ogni $g \in G$. Se G è finito, quest'ultima condizione è soddisfatta automaticamente, poiché tutti i $\chi(g)$ sono radici dell'unità. Sull'insieme dei caratteri G^* si ha una struttura di gruppo abeliano con moltiplicazione $\chi_1\chi_2(g)=\chi_1(g)\chi_2(g)$.

Fissiamo $g \in G$ e consideriamo l'applicazione $\beta(g): G^* \to \mathbf{C}^\cdot$, tale che $\chi \mapsto \chi(g)$. Si vede facilmente che questa è un carattere di G^*, cioè appartiene a G^{**}. L'applicazione risultante $\beta: G \to G^{**}$ è un omomorfismo di gruppi. Come nel caso della dualità lineare, si ha: L'applicazione canonica $\beta: G \to G^{**}$ è una immer-

sione se e solo se G è un gruppo abeliano, ed è un isomorfismo se e solo se G è un gruppo abeliano finito.

L'immagine di β è sempre un fattore abeliano massimale di G^{**}, perciò nelle costruzioni di dualità dei gruppi non commutativi occorre utilizzare nuovi principî. D'altra parte β è un isomorfismo anche per la piú ampia classe dei gruppi abeliani topologici localmente compatti, se si considerano solo i caratteri continui. Questo è l'oggetto della teoria della dualità di Pontrjagin (cfr. § 3).

L'affermazione sull'isomorfismo β si dimostra facilmente, tenendo conto del fatto che un gruppo abeliano finito si decompone nella somma diretta di sottogruppi ciclici. Supponiamo dapprima che G sia ciclico di ordine d con generatore g. Allora $\chi \in G^*$ è definito univocamente dalla radice dell'unità $\chi(g)$, che può essere di un ordine qualsiasi che divide d. Ciò implica che l'ordine di G^* è uguale a d, come per G. Inoltre, se $h \in G$ è un elemento diverso dall'identità, allora esiste sempre un $\chi \in G^*$ con $\chi(h) \neq 1$ (ad esempio, un χ tale che $\chi(g)$ sia una radice primitiva dell'unità di ordine d). Perciò l'immagine di h in G^{**} è non-banale. Ciò significa che l'omomorfismo $G \to G^{**}$ ha un nucleo banale e, dal momento che gli ordini dei gruppi G e G^{**} coincidono, esso è un isomorfismo. È facile passare al caso generale, costruendo l'isomorfismo canonico $(G_1 \times G_2)^* = G_1^* \times G_2^*$: qui G_1^* è l'insieme dei caratteri di $G_1 \times G_2$, banali su G_2, e viceversa.

Sia $\varphi : G \to H$ un omomorfismo di due gruppi abeliani finiti. Si definisce a partire da questo l'omomorfismo coniugato $\varphi^* : H^* \to G^*$, che al carattere $\chi : H \to \mathbf{C}^\cdot$ associa il carattere $\chi \circ \varphi : G \to \mathbf{C}^\cdot$. Applicando l'operazione di coniugazione ancora una volta, si ottiene l'omomorfismo $\varphi^{**} : G^{**} \to H^{**}$. Come nel caso della dualità lineare, si ha: In seguito all'identificazione di G^{**} con G e di H^{**} con H mediante β, l'applicazione φ^{**} s'identifica con φ.

Le successioni esatte di gruppi abeliani si definiscono esattamente come le successioni esatte di spazi vettoriali. Di nuovo, come nel caso della dualità lineare, si ha: Se la successione di gruppi abeliani finiti $F \xrightarrow{\varphi} G \xrightarrow{\psi} H$ è esatta in G, allora la successione $H^* \xrightarrow{\psi^*} G^* \xrightarrow{\varphi^*} F^*$ è esatta in G^*.

Siano G, H due gruppi abeliani. L'applicazione $f : G \times H \to \mathbf{C}^\cdot$ si dice accoppiamento (a valori in $\mathbf{C}^\cdot$) se essa è un omomorfismo nel primo argomento, per un valore fissato del secondo, e viceversa. Per esempio, si ha l'accoppiamento canonico $G^* \times G \to \mathbf{C}^\cdot : \langle \chi, g \rangle \to \chi(g)$. Come nel caso lineare, si definiscono i nuclei destro e sinistro dell'accoppiamento, e gli accoppiamenti non-degeneri. In modo analogo si dimostra che l'assegnazione di un accoppiamento non-degenere $f : G \times H \to \mathbf{C}^\cdot$ di gruppi abeliani finiti è equivalente all'assegnazione di un'identificazione $G \simeq H^*$ e $H \simeq G^*$.

Seguono alcuni esempi importanti per l'algebra e la teoria dei numeri.

Dualità di Kummer.

Sia k un campo contenente le radici dell'unità di periodo $n > 1$, dove n è un numero naturale che non divide la caratteristica del campo k; e K/k un'estensione finita di Galois con gruppo abeliano G di periodo n. Si consideri nel gruppo

moltiplicativo $k^\cdot$ del campo k il sottogruppo $L^\cdot$, formato dagli elementi che sono potenze n-esime in K. Evidentemente $(k^\cdot)^n \subseteq L^\cdot$. Si pone $H = L^\cdot/(k^\cdot)^n$ e si definisce l'applicazione $G \times H \to k^\cdot$ mediante la formula $\langle g, a(k^\cdot)^n \rangle \mapsto g\sqrt[n]{a}/\sqrt[n]{a}$. In altre parole, si estrae da a la radice di grado n nel campo K, e si osserva per quale radice dell'unità si moltiplica il suo automorfismo $g \in G$ del gruppo di Galois del campo K. Si dimostra facilmente che quest'applicazione è un accoppiamento (a valori in $k^\cdot$). Un aspetto piú sottile è il fatto che esso è non-degenere e definisce la dualità di Kummer fra il gruppo di Galois G e un sottofattore del gruppo moltiplicativo $k^\cdot$ del campo di base.

Questa costruzione è del tutto generale; invero essa non usa in nessun modo le proprietà specifiche del campo di base. Applicandola a campi particolari si ottengono, per esempio, l'accoppiamento di Weil (per i campi di funzioni abeliane), o il simbolo di Hilbert del residuo normato.

Residui normati.

Sia $\mathbf{Q}_p$ il campo dei numeri p-adici. Esso, come ogni campo, contiene 1, la radice dell'unità di grado 2, e il gruppo $H = \mathbf{Q}_p/\mathbf{Q}_p^{\cdot 2}$ è finito. La dualità di Kummer definisce un accoppiamento non-degenere $G \times \mathbf{Q}_p/\mathbf{Q}_p^{\cdot 2} \to \{\pm 1\}$, dove G è il gruppo di Galois dell'estensione abeliana massimale di $\mathbf{Q}_p$ di periodo 2, ottenuta aggiungendo a $\mathbf{Q}_p$ le radici quadrate di tutti i suoi elementi.

D'altra parte, si definisce l'applicazione $\mathbf{Q}_p^\cdot/\mathbf{Q}_p^{\cdot 2} \times \mathbf{Q}_p^\cdot/\mathbf{Q}_p^{\cdot 2} \to \{\pm 1\}$, ponendo per $a, b \in \mathbf{Q}_p^\cdot$

$$(a\mathbf{Q}_p^{\cdot 2}, b\mathbf{Q}_p^{\cdot 2}) = (a,b)_p = \begin{cases} +1 & \text{se } ax^2 + by^2 - z^2 = 0 \text{ per alcuni } x, y, z \in \mathbf{Q}_p, \\ & \text{non tutti nulli;} \\ -1 & \text{altrimenti.} \end{cases}$$

Questo è il simbolo di Hilbert, che costituisce un accoppiamento non-degenere. Esso è contemporaneamente simmetrico e antisimmetrico: questi due concetti coincidono per accoppiamenti a valori nei gruppi di periodo 2.

In tal modo deve esistere un isomorfismo $\mathbf{Q}_p^\cdot/\mathbf{Q}_p^{\cdot 2} \to G$, dove G è il gruppo abeliano di Galois, introdotto sopra, della estensione di $\mathbf{Q}_p$. In effetti secondo Kummer $G \simeq (\mathbf{Q}_p^\cdot/\mathbf{Q}_p^{\cdot 2})^*$, e secondo Hilbert $\mathbf{Q}_p^\cdot/\mathbf{Q}_p^{\cdot 2} \simeq (\mathbf{Q}_p^\cdot/\mathbf{Q}_p^{\cdot 2})^*$.

Tale isomorfismo effettivamente esiste e può essere costruito indipendentemente in una situazione molto piú generale. Invero, sia k un'estensione finita del campo $\mathbf{Q}_p$, G_k il gruppo di Galois dell'estensione abeliana massimale di questo campo. Allora esiste un isomorfismo $\alpha : k^\cdot \to G_k$ che è un'immersione con immagine densa, e «quasi» un isomorfismo di gruppi abeliani topologici, nel senso che sui gruppi quoziente finiti esso induce un isomorfismo. Piú precisamente, se $K \supset k$ è un'estensione abeliana finita e $H^K \subset G_k$ è il sottogruppo degli automorfismi banali su K, allora α induce un isomorfismo $\alpha_K : k^\cdot/N_{K/k}(K^\cdot) \simeq G_k/H^K$ (= gruppo di Galois di K/k). Questo è l'isomorfismo della teoria locale dei corpi di classe.

Se lo stesso k contiene le radici dell'unità di ordine n, e K/k è un'estensione abeliana massimale di periodo n, con gruppo di Galois G, allora $N_{K/k}(K^\cdot) = (k^\cdot)^n$, e α_K induce un isomorfismo di $k^\cdot/(k^\cdot)^n$ con G, cioè con il gruppo duale di

$k^{\cdot}/(k^{\cdot})^n$ secondo Kummer. In tal modo la teoria dei corpi di classe realizza la sintesi di un certo oggetto con il suo duale e perciò, come è consueto in tali casi, è altamente significativa.

I tentativi di generalizzare questi risultati al caso del gruppo di Galois completo della chiusura algebrica di un campo locale k hanno condotto negli ultimi anni alla formulazione delle importanti ipotesi di Langlands. In questa teoria «non commutativa» dei corpi di classe il posto dei caratteri del gruppo di Galois è occupato dalle sue rappresentazioni lineari di dimensione finita, e il posto dei caratteri del gruppo moltiplicativo del campo dalle rappresentazioni di dimensione infinita dei gruppi di matrici su questo campo. Alcuni di questi concetti, in termini dei quali si formulano le dualità non commutative, saranno spiegati in seguito (cfr. anche il § 3).

1.4. Gruppi non abeliani finiti e dualità.

Linearizzazione. Le note costruzioni di dualità per gruppi non necessariamente commutativi si basano su oggetti che unificano la struttura di gruppo con la struttura lineare. Due tipi fondamentali di tali oggetti sono l'anello di gruppo del gruppo G e le rappresentazioni lineari del gruppo G. Nel seguito si considererà G come gruppo finito; per il caso piú generale, si veda il § 3.

Anello di gruppo.

Sia k un campo; l'anello di gruppo $k[G]$ del gruppo G è lo spazio vettoriale delle funzioni su G a valori in k con la moltiplicazione data dalla convoluzione $(a*b)(g)=\sum_{h\in G} a(h)b(h^{-1}g)$. Se si esprime una funzione $a\in k[G]$ nella forma $\sum_{g\in G} a(g)g$, immergendo G in $k[G]$ mediante l'identificazione $g\mapsto\delta_g$ (dove $\delta_g(h)=1$ per $h=g$, e 0 per $h\neq g$), allora la convoluzione è compatibile con la moltiplicazione in G: $\delta_g*\delta_h=\delta_{gh}$. In seguito si mostrerà come, in termini di anelli di gruppo, si formuli una delle varianti della dualità.

Rappresentazioni lineari.

Sia L uno spazio vettoriale sul campo k, e sia $Gl(L)$ il gruppo degli operatori lineari invertibili $L\to L$. Si dice rappresentazione del gruppo G sul campo k un omomorfismo $T: G\to Gl(L)$; la dimensione di L si dice dimensione della rappresentazione.

Siano $T_i: G\to Gl(L_i)$, $i=1,2$, due rappresentazioni. Si dice morfismo di T_1 in T_2 un'applicazione lineare $\varphi: L_1\to L_2$ tale che $\varphi(T_1(g)l)=T_2(g)\varphi(l)$ per $l\in L_1$ e $g\in G$ arbitrari (cioè un'applicazione compatibile con le azioni di G su L_1 e L_2). Due rappresentazioni T_1 e T_2 sono equivalenti o isomorfe se esiste un morfismo di T_1 in T_2 che sia un isomorfismo lineare. Si dice somma diretta delle rappresentazioni T_1 e T_2 la rappresentazione $T=T_1\oplus T_2$ in $L_1\oplus L_2$ per la quale $T(g)(l_1,l_2)=(T(g)l_1,T(g)l_2)$ per $l_1\in L_1$, $l_2\in L_2$. In modo analogo si definisce il prodotto tensoriale $T=T_1\otimes T_2$ in $L_1\otimes L_2$. Se T è una rappresentazione di G nello spazio L, e $M\subset L$ è un sottospazio G-invariante, cioè un sottospazio tale che $T(g)m\in M$ per ogni $m\in M$, $g\in G$, allora la restrizione di T ad $M: g\mapsto T(g)|_M$ si dice sottorappresentazione di T. In modo evidente si definisce anche la rappresentazione quoziente di T su L/M. Se T non ammette sottorappresentazioni non-banali, essa si dice (algebricamente) irriducibile.

Ad ogni carattere $\chi: G\to \mathbf{C}^{\cdot}$ corrisponde una rappresentazione unidimensionale (e perciò irriducibile) di G in $\mathbf{C}$: $g\mapsto$ moltiplicazione per $\chi(g)$. Inversamente, ogni rappresentazione unidimensionale di G definisce un omomorfismo $\chi: G\to\mathbf{C}^{\cdot}$ dipendente solo dalla classe di questa rappresentazione, a meno di isomorfismi. Essa sarà senz'altro un carattere se G è finito. In tal modo G^* è identificato canonicamente con l'insieme delle classi di rappresentazioni unidimensionali del gruppo G. Questa osservazione è a fondamento di quella variante della dualità nella quale, come oggetto duale al gruppo G, si considera l'insieme $\hat{G}$ di tutte le possibili classi di rappresentazioni irriducibili del gruppo G: per ulteriori dettagli si veda il § 3.

Algebre, coalgebre e bialgebre. Uno spazio vettoriale L su un campo k si dice k-algebra (centrale) se su di esso è data una struttura di anello (non necessariamente commutativo) con unità 1, e se l'applicazione $k\to L: a\mapsto a\cdot 1$ è un omomorfismo di anelli, la cui immagine commuta con tutto L. La moltiplicazione in una k-algebra L può essere considerata come un'applicazione lineare $m: L\otimes L\to L$, $m(l_1\otimes l_2)=l_1l_2$. Tutti gli assiomi di k-algebra possono essere formulati in termini di commutatività di diagrammi di applicazioni lineari.

Uno spazio vettoriale L su k, con una data applicazione lineare strutturale (comoltiplicazione) $\mu: L\to L\otimes L$ si dice coalgebra se μ soddisfa gli assiomi ottenuti dagli assiomi per m invertendo tutte le frecce. Un'altra possibile definizione è la seguente: ((L,μ) è una coalgebra se (L^*,μ^*) è un'algebra con moltiplicazione $\mu^*: (L\otimes L)^*=L^*\otimes L^*\to L^*$). I morfismi di algebre e coalgebre, ed i loro prodotti tensoriali, si definiscono in modo evidente.

Infine, uno spazio vettoriale L dotato di moltiplicazione m e di comoltiplicazione μ si dice bialgebra se (L,m) è un'algebra, (L,μ) una coalgebra, $\mu: L\to L\otimes L$ un morfismo di algebre, e $m: L\otimes L\to L$ un morfismo di coalgebre.

La simmetria completa di tutte le condizioni su m, μ ed il loro carattere (multi)lineare permettono di verificare automaticamente i seguenti principî di dualità: se $\mathbf{L}=(L,\mu,m)$ è una bialgebra a dimensione finita, allora anche $\mathbf{L}^*=(L^*,m^*,\mu^*)$ è una bialgebra a dimensione finita; si ha un isomorfismo canonico $\mathbf{L}=\mathbf{L}^{**}$; per ogni morfismo di bialgebre $\varphi: \mathbf{L}\to\mathbf{M}$ è univocamente definito un morfismo $\varphi^*: \mathbf{M}^*\to\mathbf{L}^*$, tale che $(\varphi\circ\psi)^*=\psi^*\circ\varphi^*$ e $\varphi^{**}=\varphi$.

Il ruolo di queste costruzioni nella teoria dei gruppi finiti è definito dal fatto che l'anello di gruppo $k[G]$ di un gruppo arbitrario G è una bialgebra con la moltiplicazione definita sopra, e la comoltiplicazione $\mu(a)(g,h)=a(g)$ se $g=h$, zero altrimenti.

Dal momento che in $k[G]$ si ha una base particolare $\{\delta_g\}$, possiamo identificare canonicamente lo spazio $k[G]^*$ con $k[G]$; ciò definisce su $k[G]$ una coppia duale di moltiplicazione e comoltiplicazione. Queste saranno la moltiplicazione (puntuale) degli elementi di $k[G]$ come funzioni su G e la comoltiplicazione

$\mu(a)(g,h)=a(g,h)$. In altre parole, l'origine della convoluzione in $k[G]$ è la seguente: la moltiplicazione nel gruppo $G\times G\to G$ definisce un'applicazione degli spazi di funzioni $k[G]\to k[G\times G]=k[G]\otimes k[G]$, dopo di che questa viene dualizzata e $k[G]^*$ s'identifica con $k[G]$.

Qui, invece di $k[G]$, si sarebbe potuto prendere un altro funtore controvariante dalla categoria dei gruppi nella categoria degli anelli (cfr. § 2), per esempio l'anello di coomologia per un gruppo topologico G. In tale contesto questa costruzione è stata considerata per la prima volta da Hopf, cosicché le bialgebre e le loro varianti si chiamano anche algebre di Hopf.

Se G è un gruppo abeliano finito, si possono identificare canonicamente le bialgebre $k[G]^*$ e $k[G^*]$. Questa identificazione si effettua mediante la seguente trasformazione di Fourier: $a\mapsto\hat{a}:\hat{a}(\chi)=\sum_{g\in G}a(g)\chi(g)$. Essa è stata scoperta per la prima volta per i gruppi topologici $G=\mathbf{C}_1$ e $G=\mathbf{R}$ (serie e integrali di Fourier), ed ha, in questa variante classica, numerose applicazioni (cfr. § 3). Per i gruppi finiti $G=\mathbf{Z}/m\mathbf{Z}$ essa conduce a considerare le somme trigonometriche $\sum_{g \bmod m} a(g)e^{2\pi ig/m}$, che dai tempi di Gauss fino ai nostri giorni sono oggetto di studio in teoria dei numeri e rivelano sempre nuove proprietà profonde. Gauss ha calcolato tali somme per m primo e $a(g)=\left(\frac{g}{m}\right)=\pm 1$ in dipendenza del fatto che g sia un quadrato mod m o no, e si è servito di questi risultati per dimostrare la legge di reciprocità. In seguito (Davenport, Hasse, Weil, Serre, Deligne) si è chiarito che queste somme sono connesse con le rappresentazioni del gruppo di Galois del campo dei numeri algebrici e con le coomologie astratte delle varietà algebriche.

Se G è finito, ma non abeliano, allora la bialgebra $k[G]^*$ non è isomorfa alla bialgebra di gruppo di nessun gruppo. La categoria delle bialgebre a dimensione finita è sostanzialmente piú grande della categoria delle bialgebre di gruppo: in effetti essa è tanto grande da essere chiusa rispetto al passaggio agli oggetti linearmente duali. La struttura dei suoi oggetti «spuri», a volte chiamati anelli di gruppo generali, con moltiplicazione e comoltiplicazione non commutative, è molto poco conosciuta.

Dualità di Tanaka-Krein. Nella costruzione che ora esporremo, il ruolo di oggetto duale di un gruppo finito G è assunto dalla categoria $R[G]$ di tutte le rappresentazioni a dimensione finita di G sul campo $\mathbf{C}$. La seconda dualizzazione, che permette di riottenere G a partire da $R(G)$, si basa sul concetto di rappresentazione della categoria $R(G)$.

Si dice rappresentazione di $R(G)$ una funzione φ che ad ogni oggetto di $R(G)$, cioè ad ogni rappresentazione $T: G\to Gl(L)$ associa un operatore lineare invertibile $\varphi(T): L\to L$ con le seguenti proprietà:

a) Se $f: T_1\to T_2$ è un qualsiasi morfismo di rappresentazioni, allora $f\circ\varphi(T_1)=\varphi(T_2)\circ f$.

b) $f(T_1\otimes T_2)=f(T_1)\otimes f(T_2)$.

La classe $R(G)^*$ delle rappresentazioni della categoria $R(G)$ può essere dotata della moltiplicazione $\varphi_1\varphi_2(T)=\varphi_1(T)\varphi_2(T)$.

Ogni elemento $g\in G$ definisce una rappresentazione $\beta(g)$ della categoria $R(G)$ mediante la formula $\beta(g)T=T(g)$. L'applicazione ottenuta $\beta: G\to R(G)^*$ è un isomorfismo di gruppi (teorema di Tanaka). Si descriverà ora brevemente come si estende questo risultato ad una classe piú generale di gruppi topologici, e come si ottiene una descrizione astratta di tutte le categorie $R(G)$ corrispondenti a questi gruppi.

I gruppi finiti sono inclusi nella piú ampia categoria dei gruppi topologici compatti (cfr. § 3). La categoria $R(G)$ per un gruppo compatto G si definisce come la categoria delle rappresentazioni complesse continue a dimensione finita di G. Il teorema di Tanaka sulla ricostruzione di G a partire da $R(G)^*$ si trasporta letteralmente a questo caso piú generale. Inoltre appare la possibilità di definire in modo invariante la classe di categorie del tipo $R(G)$.

Invero Krein ha dimostrato che la categoria R degli spazi vettoriali a dimensione finita, chiusa rispetto ai prodotti tensoriali (di spazi e di morfismi) e al passaggio allo spazio (morfismo) duale, è equivalente ad $R(G)$ per un opportuno gruppo compatto G se e solo se:

a) Si ha un unico (a meno di isomorfismi) oggetto $L_0\in R$ e un isomorfismo di funtori $L_0\otimes Id\to Id$.

b) Ogni oggetto R si decompone nella somma di oggetti minimali.

c) Se L_1 e L_2 sono oggetti minimali, allora $\mathrm{Hom}_R(L_1, L_2)$ risulta o banale o uno spazio unidimensionale e, in quest'ultimo caso, L_1 e L_2 sono R-isomorfi.

Negli ultimi tempi questo risultato è stato studiato in fisica teorica (Doplicher, Haag, Roberts). In diverse teorie dei campi svolge un ruolo importante il gruppo di *gauge* G. La scelta di questo nei casi concreti è condizionata in particolare dalle cosiddette regole di superselezione, che dànno informazione proprio sull'oggetto duale a G: $R(G)$.

2. *Aspetti categoriali della dualità.*

Come si è visto nel § 1, l'oggetto duale a uno spazio vettoriale o a un gruppo abeliano consiste delle applicazioni di questo spazio (gruppo) in un certo oggetto fissato.

Il punto di vista secondo il quale l'attenzione si sposta dalla struttura interna di un oggetto alla classe dei suoi morfismi verso altri oggetti viene detto categoriale, in contrapposizione a strutturale.

Padroneggiando il linguaggio categoriale, possiamo esaminare un'intera gerarchia di dualità, a cominciare dai piú semplici effetti di «inversione delle frecce» per finire con la poco formale complementarità «descrizione strutturale/categoriale degli oggetti matematici».

Già da tempo era stato notato che coppie di costruzioni come i limiti pro-

iettivi ed induttivi si trovano in una certa relazione di dualità: ciò si rifletteva nella vecchia terminologia «spettro diretto» e «spettro inverso». Nella formulazione del linguaggio categoriale è diventato chiaro che, accanto a qualunque costruzione di un oggetto di una categoria mediante un diagramma di altri oggetti, è utile considerare la costruzione duale ottenuta per inversione di tutte le frecce nella definizione. Nonostante l'estrema semplicità di quest'idea, essa ha condotto rapidamente sia all'introduzione di nuovi concetti che alla migliore comprensione di vecchi. Per esempio, l'abbondanza di prodotti tensoriali di anelli nell'algebra commutativa degli ultimi decenni si spiega completamente con il fatto che l'inversione delle frecce (cioè il passaggio alla categoria duale degli spettri) trasforma il prodotto tensoriale nel prodotto diretto (o in quello fibrato). Il ruolo stesso dei prodotti diretti in un'arbitraria categoria diventa chiaro in analogia con il loro ruolo nella categoria degli insiemi dove i grafici delle funzioni, le leggi di composizione e molte altre strutture sono definiti in termini di prodotti.

Seguendo l'azione dell'inversione delle frecce su una categoria globalmente, e non soltanto sui diagrammi singoli, si giunge alla questione delle connessioni della categoria con quella ad essa duale. La categoria duale C^o può essere equivalente a quella iniziale C: tale è il caso della dualità degli spazi lineari a dimensione finita, o della dualità di Pontrjagin. La categoria duale può acquistare proprietà «geometriche» inaspettate, come accade con varie costruzioni degli spettri di anelli. Infine la categoria duale può ammettere una realizzazione indipendente, come nell'esempio classico della teoria di Galois.

Il funtore Hom sul prodotto di una categoria e della sua duale, a valori nella categoria degli insiemi, è formalmente simile al prodotto scalare canonico $L \times L^* \to k$ (L è uno spazio lineare su k). Estendendo questa analogia, Kan ha osservato che molti noti funtori si raggruppano in coppie i cui membri sono nella stessa relazione reciproca rispetto a Hom in cui si trovano le coppie di operatori lineari coniugati. Ciò ha condotto ad una trattazione categoriale unificata di dualità differenti quali la dualità di Frobenius nella teoria delle rappresentazioni dei gruppi, e la dualità «sospensione/spazio dei cappi» in topologia omotopica.

Infine il confronto del linguaggio categoriale con quello strutturale rivela in essi le caratteristiche di una reciproca complementarità, tanto accentuata che gli specialisti di fondamenti della matematica considerano ora la possibilità di una costruzione alternativa dei fondamenti sulla base delle categorie, e non degli insiemi con struttura. Indipendentemente da ciò, per il topologo o per il geometra algebrico è già divenuta attuale la padronanza degli aspetti operativi di tale complementarità: le regole d'introduzione di una struttura su di un oggetto di una categoria, o il considerare, invece di un oggetto, il funtore che rappresenta questo oggetto.

Questo paragrafo è dedicato al chiarimento di tutte queste idee per mezzo di esempi molto semplici. Per la definizione generale di categorie e di funtori, si rimanda il lettore al § 6 dell'articolo «Applicazioni» di questa stessa *Enciclopedia*. Lí sono anche portati i primi esempi di categorie e di costruzioni categoriali, e nel § 4 è contenuta la descrizione di una serie di strutture matematiche e di loro applicazioni.

2.1. Inversione delle frecce.

Sia C una categoria. Si ricordi che la categoria duale C^o ha gli stessi oggetti di C: $\mathrm{Ob}\,C = \mathrm{Ob}\,C^o$. Si indichi con $X^o \in \mathrm{Ob}\,C^o$ l'oggetto corrispondente a $X \in \mathrm{Ob}\,C$. Per definizione, $\mathrm{Hom}_{C^o}(X^o, Y^o) = \mathrm{Hom}_C(Y, X)$ e $(fg)^o = g^o f^o$, dove f e g sono morfismi in C e f^o, g^o i corrispondenti morfismi in C^o. In altre parole, C^o si ottiene da C per inversione di tutte le frecce.

Costruzioni duali.

Supponiamo che in C sia definita una costruzione, applicabile anche in C^o. Allora essa, trasportata in C^o, e considerata poi come costruzione su C, si chiama costruzione duale a quella iniziale.

Esempio. Si abbia in C un diagramma commutativo (quadrato) di morfismi

$$\begin{tikzcd} X \arrow[d, "\varphi"'] & Z \arrow[l, "u"'] \arrow[d, "v"] \\ S & Y \arrow[l, "\psi"] \end{tikzcd}$$

Tale quadrato si dice cartesiano se per ogni oggetto su S: $T \to S$ l'applicazione $\mathrm{Hom}_S(T, Z) \to \mathrm{Hom}_S(T, X) \times \mathrm{Hom}_S(T, Y) : g \mapsto (u \circ g, v \circ g)$ è biiettiva. Qui Hom_S sono i morfismi nella categoria C_S degli oggetti su S. In tal caso gli oggetti Z e u, v sono definiti da $X \xrightarrow{\varphi} S \xleftarrow{\psi} Y$ univocamente a meno di isomorfismi, e Z viene detto prodotto fibrato (su S) degli oggetti X, Y: $Z = X \times_S Y$.

La costruzione duale parte dal quadrato cartesiano

$$\begin{tikzcd} X \arrow[r, "u"] & Z \\ S \arrow[u, "\varphi"] \arrow[r, "\psi"'] & Y \arrow[u, "v"'] \end{tikzcd}$$

e definisce la somma disgiunta $Z = X \sqcup_S Y$. Essa può essere espressa nel modo piú rapido nella forma $X \sqcup_S Y = (X^o \times_{S^o} Y^o)$, utilizzando l'identificazione naturale $(C^o)^o = C$.

Ad esempio, nella categoria degli anelli commutativi con unità Ann, si ha $A \sqcup_{\mathbf{C}} B = A \otimes_{\mathbf{C}} B$ (prodotto tensoriale di anelli su $\mathbf{C}$). La categoria $\mathbf{Ann}^o$ si chiama categoria degli schemi affini. Invece di A^o si usa scrivere $\mathrm{Spec}\,A$. In queste notazioni $\mathrm{Spec}\,A \times_{\mathrm{Spec}\,\mathbf{C}} \mathrm{Spec}\,B = \mathrm{Spec}(A \otimes_{\mathbf{C}} B)$.

Esempio. Sia $\bar{X}: \ldots \to X_{-1} \xrightarrow{u_{-1}} X_0 \xrightarrow{u_0} X_1 \xrightarrow{u_1} \ldots \to X_n \xrightarrow{u_n} \ldots$ una successione di oggetti e morfismi nella categoria C. Si chiama morfismo di $\bar{X}$ nell'oggetto $X \in \operatorname{Ob} C$ la successione di morfismi $v_i: X_i \to X$ tali che tutti i diagrammi

$$\begin{tikzcd} X_i \arrow[rr, "u_i"] \arrow[dr, "v_i"'] & & X_{i+1} \arrow[dl, "v_{i+1}"] \\ & X & \end{tikzcd}$$

siano commutativi. Supponiamo che un morfismo $(v_i): \bar{X} \to X$ sia universale nel senso che per ogni altro morfismo $(w_i): \bar{X} \to Y$ si abbia un unico morfismo $w: X \to Y$ tale che il diagramma

$$\begin{tikzcd} \bar{X} \arrow[rr, "(v_i)"] \arrow[dr, "(w_i)"'] & & X \arrow[dl, "w"] \\ & Y & \end{tikzcd}$$

sia commutativo nel senso evidente della parola.

Allora (v_i) e X sono definiti univocamente a meno di isomorfismi, e X si dice limite induttivo della successione $\bar{X} = (X_i)$ nella categoria $C: X = \varinjlim X_i$.

La costruzione duale si chiama costruzione del limite proiettivo: $\varprojlim X_i = (\varinjlim X_i^{\circ})^{\circ}$.

Ad esempio, se p è un numero primo, $X_{-n} = \mathbf{Z}/p^n\mathbf{Z}$ per $n \geq 0$, $X^n = 0$ per $n \geq 0$, u_{-n} è l'applicazione di riduzione $\mathbf{Z}/p^n\mathbf{Z} \to \mathbf{Z}/p^{n-1}\mathbf{Z}$ o zero per $n \geq 0$, allora $\varprojlim X_n = \mathbf{Z}_p$ è l'anello degli interi p-adici. Si vede facilmente che $\varinjlim X_i^* = \mathbf{Q}_p/\mathbf{Z}_p$, dove X_i^* è il gruppo abeliano duale di X_i, e il sistema (X_i^*) si ottiene da (X_i) applicando la dualità dei gruppi abeliani.

Categorie autoduali.

In alcuni casi importanti le categorie C e C° risultano equivalenti. Questo concetto sostituisce il concetto piú semplice, ma meno utile, di isomorfismo di categorie. Invero siano B, C due categorie, e $B \underset{G}{\overset{F}{\rightleftarrows}} C$ due funtori. I funtori F e G definiscono una equivalenza di B e C se i funtori $G \circ F: B \to B$ e $F \circ G: C \to C$ sono isomorfi al funtore identico (per il concetto di morfismo e di isomorfismo di funtori si veda l'inizio del § 2.2).

Una categoria C si dice autoduale se le categorie C e C° sono equivalenti.

Esempio. Le categorie FAb dei gruppi abeliani finiti, e FLin degli spazi vettoriali a dimensione finita sono autoduali. L'equivalenza è realizzata dai funtori $X \mapsto (X^*)^{\circ}$ e $Y^{\circ} \mapsto Y^*$ nelle notazioni dei §§ 2 e 3. In ciò consiste il contenuto

di tutti i principî di dualità descritti nei §§ 1 e 3. Anche la categoria LCAb dei gruppi abeliani topologici localmente compatti è autoduale (cfr. § 3): in ciò consiste il principio di dualità di Pontrjagin.

Realizzazioni importanti della categoria C°.

Spesso accade che la categoria C° non sia equivalente a C, ma ammetta una costruzione indipendente da C. Alcune delle piú profonde costruzioni matematiche si interpretano in modo naturale in questi termini.

Esempio: teoria di Galois. Sia k un campo, $\bar{k}$ la sua chiusura algebrica. Per semplicità di formulazione limitiamoci al caso di caratteristica 0. Allora il gruppo degli automorfismi del campo $\bar{k}$ su k è un gruppo topologico di Galois G, e la teoria di Galois stabilisce l'equivalenza fra due categorie:

a) la categoria dei sottogruppi chiusi di G, con le inclusioni naturali quali morfismi;

b) la categoria duale dei campi K, $k \subset K \subset \bar{k}$, con le inclusioni naturali quali morfismi.

L'equivalenza si realizza mediante i funtori: (sottogruppo $H \subset G$) $\mapsto$ (campo degli elementi di $\bar{k}$, invarianti rispetto ad H) e (campo K) $\mapsto$ (gruppo degli elementi di G che agiscono come l'identità su K).

Esempio: dualità nelle algebre di Banach. Essa stabilisce l'equivalenza della categoria degli spazi di Hausdorff compatti con una certa sottocategoria delle algebre di Banach commutative: si veda la formulazione precisa nel § 3.3.

Questo esempio è una delle note concretizzazioni della dualità generale geometria/algebra o spazio/funzioni su di esso.

Un altro esempio è fornito dagli schemi affini di Grothendieck, che si riallacciano anche alla trattazione algebrica delle superfici di Riemann secondo Kronecker-Weber. Nella costruzione di Grothendieck si vede chiaramente come la dualizzazione della categoria degli anelli commutativi riveli in essi delle possibilità geometriche del tutto inaspettate. Gli schemi affini Spec A diventano semplicemente degli oggetti geometrici locali, con i quali si possono comporre schemi globali, immergendo la categoria Ann$^{\circ}$, che conserva ancora tracce della sua origine algebrica, nella categoria già completamente geometrica di tutti gli schemi. La seconda caratteristica tipica della costruzione di Grothendieck è che in essa gli oggetti primitivi diventano le funzioni, e non lo spazio su cui esse sono definite; inoltre queste funzioni non sono neppure definite univocamente dai loro valori nei punti (i nilpotenti in A si annullano in tutti i punti di Spec A). Tale dualizzazione, come strumento di generalizzazione di strutture note, è un procedimento ampiamente diffuso.

2.2. Funtori aggiunti.

Siano $F, G: C \to C'$ due funtori. Si dice morfismo funtoriale, o trasformazione naturale $\varphi: F \to G$ una famiglia di morfismi $\varphi_X: F(X) \to G(X)$ nella ca-

tegoria C', uno per ogni oggetto $X \in \mathrm{Ob}\, C$, tali che, se $X \xrightarrow{f} Y$ è un arbitrario morfismo in C, il diagramma

$$\begin{tikzcd} F(X) \arrow[r, "\varphi_x"] \arrow[d, "F(f)"'] & G(X) \arrow[d, "G(f)"] \\ F(Y) \arrow[r, "\varphi_y"'] & G(Y) \end{tikzcd}$$

è commutativo. La composizione di morfismi funtoriali si definisce in modo evidente. Il morfismo funtoriale identico, ad esempio $F \to F$, è la famiglia delle applicazioni identiche $F(X) \to F(X)$. Un morfismo funtoriale si dice isomorfismo se è dotato d'inverso destro e sinistro, cioè di quei morfismi che dànno il funtore identico per composizione con esso.

Siano, ad esempio, Y, Y' due oggetti di C, $w: Y \to Y'$ un morfismo, $h_Y, h_{Y'}$ i due funtori da C^o in Insiemi (la categoria degli insiemi) che rappresentano Y e Y'.

Allora i morfismi

$$h_Y(X) = \mathrm{Hom}_C(X, Y) \xrightarrow{\varphi_X} h_{Y'}(X) = \mathrm{Hom}_C(X, Y') : u \mapsto wu$$

uno per ogni $X \in \mathrm{Ob}\, C$, formano un morfismo funtoriale $h_w : h_Y \to h_{Y'}$. È facile vedere che $h_{w_1 w_2} = h_{w_1} h_{w_2}$.

Sia Funct (C, C') la classe dei funtori da C in C'. La si considererà come la classe degli oggetti di una categoria (con le stesse notazioni) i cui morfismi sono le trasformazioni naturali di funtori. Trascuriamo qui le questioni di teoria degli insiemi (la differenza fra insiemi e classi). Allora il risultato delle considerazioni precedenti si descrive nel modo piú semplice come la costruzione di un morfismo canonico da un'arbitraria categoria C nella categoria $\hat{C} = \mathrm{Funct}(C^o, \text{Insiemi})$. Quest'ultima si chiama categoria dei prefasci di insiemi su C. Essa si ottiene da C con una specie di duplice dualizzazione: il passaggio a C^o e poi ai funtori nella categoria universale degli insiemi.

Il funtore canonico $C \to \hat{C} : X \mapsto h_X$ è pienamente fedele (*pleinement fidèle*): l'applicazione naturale $\mathrm{Hom}_C(X, Y) \to \mathrm{Hom}_{\hat{C}}(h_X, h_Y)$ è biiettiva. In particolare, l'immagine di C in $\hat{C}$ è una sottocategoria piena di $\hat{C}$, equivalente a C.

Una descrizione un po' piú simmetrica di questa stessa situazione si ottiene se si considera l'applicazione $(X^o, Y) \mapsto \mathrm{Hom}(X, Y)$ come bifuntore canonico $C^o \times C \to$ Insiemi, che ricorda formalmente l'accoppiamento di spazi vettoriali o di gruppi. Continuando in questa analogia, si considerino una categoria B e due funtori $F: B \to C$ e $G: C \to B$. Questi funtori si dicono aggiunti l'uno all'altro se i bifuntori

$$B^o \times C \to \text{Insiemi} : (X^o, Y) \mapsto \mathrm{Hom}_C(F(X), Y) \text{ e } (X^o, Y) \mapsto \mathrm{Hom}_B(X, G(Y))$$

sono isomorfi. Nelle notazioni F è aggiunto a sinistra di G, e G è aggiunto a destra di F.

Esempio. Nella categoria Insiemi è definito il prodotto diretto e, per arbitrari insiemi X, Y, Z, esiste l'identificazione canonica $\mathrm{Hom}(X, \mathrm{Hom}(Y, Z)) = \mathrm{Hom}(X \times Y, Z)$, che all'applicazione $x \mapsto (f_x : Y \to Z)$ associa l'applicazione $\langle x, y \rangle \mapsto f_x(y)$. Verificando tutte le necessarie proprietà funtoriali, ci si può convincere del fatto che il funtore $X \mapsto X \times Y$ è aggiunto a sinistra del funtore $Z \mapsto \mathrm{Hom}(Z, Y)$.

Esempio. Analogamente, nella categoria degli spazi vettoriali a dimensione finita sono aggiunti i funtori $X \mapsto X \otimes Y$ e $Z \mapsto \mathrm{Hom}(Z, Y)$.

Esempio: dualità di Frobenius. Siano G un gruppo finito, $H \subset G$ un suo sottogruppo, $R(G)$ e $R(H)$ le corrispondenti categorie delle rappresentazioni finite sul campo k (cfr. § 1.4). Restringendo ogni rappresentazione $T: G \to Gl(L)$ al sottogruppo H, si ottiene una rappresentazione di H. Ciò definisce il funtore di restrizione res : $R(G) \to R(H)$. Inversamente, da ogni rappresentazione U del sottogruppo H nello spazio L si può ottenere la cosiddetta rappresentazione indotta ind U del gruppo G. Questa si può realizzare, ad esempio, nello spazio delle funzioni f su G a valori in L per cui $f(hg) = U(h)f(g)$ per ogni $h \in H$, $g \in G$. L'azione di G si definisce nella maniera seguente: $[\mathrm{ind}\, U(g)]f(g_1) = f(g_1 g)$. Il teorema di dualità di Frobenius afferma (nella formulazione categoriale) che il funtore res è aggiunto a sinistra del funtore ind.

Esempio. Sia Gr la categoria dei gruppi, $G: Gr \to$ Insiemi il funtore dimenticante della struttura di gruppo, che associa ad ogni gruppo il gruppo stesso come insieme. Il funtore aggiunto ad esso F: Insiemi $\to Gr$ esiste, e definisce per ogni insieme S il gruppo libero $F(S)$ generato da questo insieme. In effetti gli insiemi $\mathrm{Hom}_{Gr}(F(S), T)$ e $\mathrm{Hom}_{\text{Insiemi}}(S, G(T))$ s'identificano naturalmente per ogni gruppo T: l'omomorfismo di $F(S)$ in T è definito dall'assegnazione delle immagini dei generatori liberi da S in T, e queste immagini possono essere scelte arbitrariamente.

In genere tutte le possibili strutture algebriche libere si definiscono nel modo piú comodo proprio mediante il funtore aggiunto al funtore dimenticante della corrispondente struttura. Le costruzioni abituali degli oggetti liberi diventano allora teoremi di esistenza.

2.3. Dualità struttura/categoria.

Come si è già detto nell'articolo «Applicazioni», i punti di vista strutturale e categoriale su oggetti matematici quali i gruppi o gli spazi topologici sono complementari. Dal punto di vista strutturale un oggetto è dato come un insieme, formato da elementi, con determinate leggi di composizione fra elementi, o proprietà di elementi, relazioni, proprietà delle relazioni, ecc. Dal punto di vista categoriale un oggetto è considerato come elemento della classe di tutti gli oggetti della stessa specie, e la sua individualità è definita dalla classe di tutte le sue applicazioni in altri oggetti o di altri oggetti in esso, cioè dal funtore che esso rappresenta. In altre parole l'approccio strutturale presenta un oggetto dall'interno, quello categoriale dall'esterno. Questa complementarità possiede vari aspetti piú particolari, alcuni dei quali saranno ora elencati.

Nucleo/immagine.

Se X è un insieme dotato di struttura, Y un altro insieme con la stessa struttura e $f: X \to Y$ un morfismo suriettivo che conserva la struttura, allora f definisce su X una relazione di equivalenza di tipo particolare: le sue classi di equivalenza sono le f-controimmagini dei punti di Y. Questa relazione di equivalenza è una nuova struttura su X, che definisce f univocamente a meno di isomorfismi. Nella maggioranza dei casi si può dare una comoda descrizione strutturale di tutte le relazioni di equivalenza corrispondenti a tutte le possibili f. Ad esempio nella categoria dei gruppi tale relazione è definita dall'assegnazione dell'immagine inversa dell'identità, che può essere un qualsiasi divisore normale $Z \subseteq X$, cioè un sottogruppo di X con la proprietà $xZx^{-1} = Z$ per ogni $x \in X$. Nella categoria degli spazi vettoriali essa è definita dall'assegnazione di un arbitrario sottospazio $Z \subseteq X$. Nella categoria degli anelli commutativi essa è definita dall'assegnazione di un ideale $\mathfrak{z} \subseteq X$, l'immagine inversa dello zero. Queste immagini inverse si dicono nuclei di f e gli Y-fattori di X rispetto a questi nuclei immagini di f. In geometria algebrica gli ideali primi $\mathfrak{p}$ di un anello A si chiamano punti schematici dello spettro, e gli omomorfismi di A nei campi algebricamente chiusi punti geometrici di Spec A. Associando all'ideale $\mathfrak{p}$ un omomorfismo di A nella chiusura algebrica del campo quoziente $A/\mathfrak{p}$, e ad un punto geometrico il suo nucleo, si ottiene una certa dualità tra i punti dei due tipi. La geometria algebrica classica considerava esclusivamente punti geometrici, in quella contemporanea i punti schematici hanno un ruolo importante.

La dualità logica generale testo/significato o, in un'accezione piú particolare, linguaggio/interpretazione, ammette una descrizione analoga. Un linguaggio formale può essere considerato come una struttura (insieme dei simboli, insieme delle formule, insieme degli assiomi, insieme delle deduzioni formali). La sua interpretazione è un morfismo di questa struttura nell'insieme, o classe, definito dall'assegnazione dei valori di tutti i simboli di operazione e di relazione (predicati) nel linguaggio: il valore del simbolo di operazione è un'operazione, quello del simbolo di relazione è una relazione, ecc. La correlazione interna di questo concetto esterno di interpretazione è fornita dall'insieme delle formule del linguaggio che diventano vere nella data interpretazione. Per i linguaggi del primo ordine si possono descrivere tutti i sottoinsiemi di formule T che possono essere veri in un'appropriata interpretazione (modelli di T): questi sono i sottoinsiemi minimali di formule che contengano tutti gli assiomi logici e siano non-contraddittori e chiusi rispetto alle regole di deduzione. In ciò consiste il contenuto del teorema di completezza di Gödel che costituisce la parte principale della dualità linguaggio/interpretazione nell'ambito dei modelli formali di entrambi i membri di questa coppia.

Categoria/struttura.

Se la categoria C è piccola, cioè i suoi oggetti formano un insieme e non una classe, allora essa stessa può essere considerata una struttura, definibile mediante l'assegnazione dell'insieme $\mathrm{Ob}\, C$, degli insiemi $\mathrm{Hom}_C(X, Y)$ per ogni $X, Y \in \mathrm{Ob}\, C$, e delle leggi di composizione fra questi. Un funtore $F: C \to C'$ da una categoria piccola in un'altra è un morfismo di queste strutture. La classe delle categorie piccole, con i funtori in qualità di morfismi, forma essa stessa una categoria; se questa classe è un insieme, essa può essere considerata a sua volta come una struttura, ecc. In tal modo la contrapposizione struttura/categoria e il suo parallelismo con la contrapposizione locale/globale (in quest'ordine) non sono assoluti. Piuttosto si ha a che fare con una gerarchia di livelli di descrizione, al cui livello piú basso si trovano le descrizioni di alcune strutture fondamentali come categorie (spazio topologico, insieme ordinato, ecc.).

Oggetto/funtore.

Il passaggio dall'esame di un insieme concreto dotato di struttura, per esempio uno spazio topologico X, all'esame di questo come oggetto della categoria dei gruppi cancella l'individualità degli elementi di X. Tuttavia la successiva sostituzione di X con il funtore h_X, da esso rappresentato, permette di associare ad X il vastissimo nuovo insieme (classe) dei «punti categoriali» di X: $\bigcup_Y \mathrm{Hom}(Y, X)$, con la struttura formata dalle applicazioni derivanti dai morfismi $Y_1 \to Y_2$. Se anche inizialmente X era un oggetto di una categoria astratta, dopo l'immersione canonica $C \to \hat{C}$ esso «diventa» un insieme strutturato, dopo di che su di esso si possono trasportare le abituali costruzioni insiemistiche, per esempio introdurre la definizione di «oggetto con una data struttura»: si veda la descrizione degli oggetti gruppali di una categoria nell'articolo «Applicazioni».

Inoltre, ponendo $\hat{X} = h_X$, si ha $\mathrm{Hom}_C(X, Y) = \mathrm{Hom}_{\hat{C}}(\hat{X}, \hat{Y})$, cosicché la sostituzione di X con $\hat{X}$ riflette in modo del tutto adeguato tutte le proprietà categoriali di X. Ma la categoria $\hat{C}$ contiene oggetti di tipo piú generale che la categoria C, e può essere chiusa rispetto ad operazioni fondamentali rispetto alle quali C non è chiusa. Per esempio, nella categoria degli schemi Sch non è possibile fattorizzare rispetto a semplici relazioni di equivalenza, come l'azione di un gruppo finito. Ciò è possibile in $\mathrm{Sch}^\wedge$. Aggiungendo all'immagine di Sch in $\mathrm{Sch}^\wedge$ i risultati di tali operazioni, otteniamo un'importante generalizzazione del concetto di schema: gli «spazi algebrici» di Michael Artin.

3. *Dualità e continuità.*

3.1. Esempi classici.

Serie ed integrali di Fourier.

Sia $f: \mathbf{R} \to \mathbf{C}$ una funzione periodica a valori complessi di periodo 1: $f(x+1) = f(x)$. Il numero $\hat{f}_n = \int_0^1 f(x) e^{-2\pi i n x}$ si dice n-esimo coefficiente di Fourier di f. Se $f(x)$ è la somma d'una serie assolutamente convergente – la sua serie di Fourier – $f(x) = \sum_{n=-\infty}^{\infty} a_n e^{2\pi i n x}$, allora $a_n = \hat{f}_n$, il che giustifica il nome. Suppo-

niamo che $f(x)$ descriva un certo processo fisico periodico o una oscillazione (x è il tempo). Se tale processo è lineare, allora le funzioni $\{e^{2\pi inx}\}$ possono descrivere un sistema di oscillazioni fondamentali (armoniche). La serie di Fourier esprime una qualsiasi oscillazione come combinazione di oscillazioni fondamentali, e $\hat{f}_n$ rappresenta l'intensità della n-esima armonica.

Nel caso non-periodico, se all'infinito la funzione $f: \mathbf{R} \to \mathbf{C}$ decresce abbastanza rapidamente, essa può ancora essere rappresentata come somma di armoniche, ma con una distribuzione continua di frequenze. L'intensità dell'armonica di frequenza p è

$$\hat{f}(p) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) e^{-2\pi ipx} dx$$

e la funzione $f(x)$ si ricostruisce a partire da $\hat{f}(p)$ per mezzo dell'integrale di Fourier

$$f(x) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \hat{f}(p) e^{2\pi ipx} dp.$$

Per esempio, $f(x)$ può essere la funzione d'onda di una particella nella rappresentazione delle coordinate, e $\hat{f}(p)$ in quella degli impulsi (cfr. § 3.3). La localizzazione di $f(x)$ come pacchetto d'onda spazialmente compatto si ottiene a prezzo di una dispersione della regione degli impulsi p, con grandi valori di $\hat{f}(p)$.

Nel caso di funzioni su $\mathbf{R}^n$ le definizioni sono analoghe, ma $e^{2\pi ipx}$ è sostituito da $e^{2\pi i(p_1x_1+\dots+p_nx_n)}$, dove $(p_1, \dots, p_n)$ è un vettore in $\mathbf{R}^{n*}$, lo spazio duale.

Il passaggio dalla funzione f a $\{\hat{f}_n\}$ o a $\{\hat{f}(p)\}$ si dice trasformazione di Fourier. Una funzione periodica può essere considerata una funzione sulla circonferenza unitaria, che è un gruppo topologico compatto. Allora $\{\hat{f}_n\}$ è una funzione sull'insieme dei suoi caratteri $\{e^{2\pi inx}\}$, isomorfo al gruppo discreto dei numeri interi. Analogamente, la trasformata di Fourier di una funzione $f(x)$ su $\mathbf{R}$ è una funzione $\hat{f}(p)$ sul gruppo dei caratteri unitari del gruppo $\mathbf{R}$: $\{x \mapsto e^{2\pi ipx}\}$, $p \in \mathbf{R}$.

Storicamente la serie e l'integrale di Fourier sono stati il punto di partenza per molte costruzioni fondamentali: qui ci si limiterà a considerarle come modelli della teoria della dualità in una situazione topologica.

Trasformazione di Radon.

Sia E uno spazio euclideo, $\check{E}$ l'insieme degli iperpiani in esso. La trasformazione di Radon trasforma una funzione f su E in una funzione $\check{f}$ su $\check{E}$: $\check{f}(P) = \int_P f d\mu_P$, dove μ_P è la misura euclidea su P. La funzione f si può ricostruire a partire da $\check{f}$: $f(x)$ è una certa media della derivata seconda di $\check{f}(P)$ (rispetto a tutti i $P \ni x$) nella direzione perpendicolare a P, se E ha dimensione dispari. Dal momento che le superfici di livello del carattere $e^{2\pi i\langle p, x\rangle}$, $x \in E$, $p \in E^*$, che compare nella formula per la trasformata di Fourier di f, sono gli iperpiani $\langle p, x\rangle = \text{cost}$, la trasformazione di Radon può spesso essere considerata un'operazione piú profonda. A livello di funzioni sullo spazio essa è una immediata generalizzazione dell'idea classica di dualità proiettiva.

Distribuzioni di Schwartz.

Sia $\mathfrak{D}$ lo spazio delle funzioni complesse infinitamente differenziabili sullo spazio reale n-dimensionale, ciascuna delle quali è uguale a zero all'esterno di una sfera di raggio sufficientemente grande. Nello spazio $\mathfrak{D}$ è assegnato un concetto di convergenza: una successione (φ_j) di funzioni di $\mathfrak{D}$ converge a zero se tutte le φ_j si annullano al di fuori di una sfera fissata e se le loro derivate di ogni ordine $(\varphi_j^{(K_1 \dots K_n)})$ tendono a zero uniformemente.

Cosí come ogni vettore in uno spazio vettoriale a dimensione finita può essere considerato un funzionale (sullo spazio duale), le funzioni sommabili su $\mathbf{R}^n$ possono essere considerate funzionali su $\mathfrak{D}$. A tale fine si associa alla funzione f l'applicazione lineare $T_f: \mathfrak{D} \to \mathbf{C}$ secondo la formula $T_f(g) = \int_{\mathbf{R}^n} f(x_1, \dots, x_n) g(x_1, \dots, x_n) dx_1, \dots, dx_n$. Questa applicazione è continua in g.

In generale ogni funzionale lineare continuo $T: \mathfrak{D} \to \mathbf{C}$ si dice distribuzione o funzione generalizzata. Le distribuzioni formano uno spazio vettoriale $\mathfrak{D}'$ contenuto nello spazio $\mathfrak{D}^*$, duale algebrico di $\mathfrak{D}$. In esso è contenuto lo spazio dei funzionali del tipo T_f, ma tali funzionali non esauriscono affatto $\mathfrak{D}'$. Un noto esempio di funzione generalizzata è la funzione delta di Dirac δ_a, $a \in \mathbf{R}^n$, per la quale $\delta_a(f) = f(a)$ per ogni $f \in \mathfrak{D}$. Anche il funzionale «$f \mapsto$ valore della derivata di f secondo una certa direzione in un dato punto» è una funzione generalizzata.

Questa definizione si giustifica con le seguenti considerazioni fisiche. Ci si può rappresentare una funzione continua f su $\mathbf{R}^n$ come densità di distribuzione di una certa grandezza (massa, carica, ecc.). Ciò significa che la quantità di questa grandezza nella regione $U \subset \mathbf{R}^n$ è uguale a $\int_U f dx$. La funzione generalizzata δ_a rappresenta allora l'idealizzazione di una carica (o massa) isolata concentrata nell'unico punto a. La densità è allora infinita in questo punto e si annulla identicamente negli altri, cosicché essa non può essere rappresentata da una funzione sommabile. Analogamente, il funzionale «valore della derivata in un punto» può descrivere, ad esempio, un dipolo isolato: il limite di un sistema di due cariche di grandezza ε e $-\varepsilon$ alla distanza $1/\varepsilon$, quando le cariche si avvicinano nella direzione del vettore che individua la direzione di differenziazione.

Caratterizzare una distribuzione come funzionale continuo sullo spazio $\mathfrak{D}$ delle «funzioni di prova» corrisponde ad una delle idealizzazioni del processo di misura macroscopica in fisica: noi non siamo mai in grado di misurare la densità di una grandezza in un punto, ma solo un suo valor medio in una regione, idealmente piccola ad arbitrio. Questo valor medio con «funzione peso» $f \in \mathfrak{D}$ è il risultato di un'unica misura, mentre l'intera famiglia di medie rispetto alle $f \in \mathfrak{D}$, cioè il funzionale completo su $\mathfrak{D}$, è la massima informazione teoricamente possibile sulla distribuzione. Ne segue che le medie possono esistere anche se non esiste alcuna densità, per esempio se la distribuzione ha una struttura fortemente granulare.

Quest'idea di media ha un carattere molto generale. Essa viene usata per esempio nella teoria dei processi stocastici e nella teoria quantistica dei campi,

dove gli osservabili locali dei campi sono solo operatori generalizzati. Non è possibile qui soffermarsi piú a lungo su questo punto.

Commento. Si è mostrato brevemente che le piú semplici idee algebriche di dualità (gruppi abeliani e spazi vettoriali) conducono, nell'ambito dell'analisi, ai modelli matematici di importanti complementarità quali onda/particella, discreto/continuo, locale/globale, ecc.

Il ruolo di queste fondamentali idee algebriche non deve essere messo in ombra dal fatto che in tutte le costruzioni concrete dell'analisi i problemi della scelta dello spazio adatto, della sua topologia, la dimostrazione delle stime necessarie, ecc., occupano uno spazio incomparabilmente piú ampio della semplice algebra. Nel seguito si esporranno un piccolo numero di esempi caratteristici, tratti da alcuni importanti capitoli dell'analisi funzionale, che utilizzano in vari modi l'idea di dualità.

3.2. Dualità lineare.

Spazi vettoriali topologici.

Modi generali per assegnare una topologia.

Nell'articolo «Applicazioni» si è spiegato che dare una topologia su un insieme X significa assegnare una famiglia di sottoinsiemi di questo, detti aperti. Tale famiglia deve contenere $\varnothing$ e X, ed essere chiusa rispetto a intersezioni finite e unioni arbitrarie. Un insieme dotato di una topologia si chiama spazio topologico.

Occorre ancora introdurre alcune definizioni di uso generale, necessarie per il seguito. I complementari degli insiemi aperti si dicono chiusi. Il piú piccolo insieme chiuso contenente il sottoinsieme $Y \subseteq X$ si dice *chiusura* di Y. Un insieme Y contenente $x \in X$ e un aperto $U \ni x$ si dice intorno di x.

Sia B una sottofamiglia di insiemi aperti di X con la seguente proprietà: per ogni intorno Y di un qualsiasi punto $x \in X$ si può trovare un $U \in B$ tale che $Y \supseteq U \ni x$. Una tale famiglia B si dice sistema fondamentale di intorni, o base della topologia di X. L'intera topologia di X si ricostruisce univocamente a partire da una sua qualsiasi base. Per esempio, se X, Y sono spazi topologici, si può definire univocamente la topologia prodotto su $X \times Y$, nella quale serviranno da base i prodotti $U \times V$, dove $U \subseteq X$, $V \subseteq Y$ sono insiemi aperti arbitrari.

Un'applicazione tra spazi topologici $f: X \to Y$ si dice continua se la f-controimmagine di un qualsiasi insieme aperto di Y è un aperto di X.

Uno spazio topologico Y si dice compatto se da ogni suo ricoprimento con insiemi aperti si può estrarre un sottoricoprimento finito, e discreto se ogni suo sottoinsieme è aperto (e perciò anche chiuso).

Spazi vettoriali topologici.

Sia K uno dei campi $\mathbf{R}$ o $\mathbf{C}$. Si dice spazio vettoriale topologico (SVT) su K uno spazio vettoriale L su K, su cui sia data una topologia tale che le applicazioni

$L \times L \to L : \langle l_1, l_2 \rangle \mapsto l_1 + l_2$, e $K \times L \to L : \langle a, l \rangle \mapsto al$ siano continue. Si possono considerare come esempio gli spazi a dimensione finita K^n con la topologia prodotto: questa è l'unica topologia che rende K^n uno SVT. Nel seguito di questo capitolo c'interesseremo principalmente agli SVT a dimensione infinita.

I casi $K = \mathbf{R}$ e $\mathbf{C}$ sono gli unici considerati nell'analisi funzionale classica. Negli ultimi decenni hanno richiamato un'attenzione sempre crescente gli SVT su campi topologici quali i numeri p-adici $\mathbf{Q}_p$ o il completamento $\mathbf{C}_p$, di una chiusura di $\mathbf{Q}_p$. La costruzione dell'analisi funzionale su $\mathbf{C}_p$, che tenga conto dell'azione del gruppo di base, probabilmente avrà applicazioni essenziali in teoria dei numeri.

Gli SVT concreti a infinite dimensioni si costruiscono spesso per stadi. Sia X un insieme. Il piú semplice spazio ad esso collegato è lo spazio di tutte le funzioni su X a valori in $\mathbf{R}$ o $\mathbf{C}$; esso è a dimensione infinita se X è infinito. In genere esso è eccessivamente grande, e se ne estrae un sottospazio di funzioni soddisfacenti condizioni supplementari. Quando X è uno spazio topologico, si possono considerare le funzioni continue; quando X è una varietà differenziabile, quelle differenziabili, ecc. Altri tipi di limitazione saranno considerati in seguito. La non-univocità della scelta degli spazi funzionali associati a un dato X, e della topologia su di essi, è il motivo principale delle complicazioni rispetto al caso a dimensione finita.

La topologia su un dato spazio di funzioni spesso si assegna in base alla richiesta che alcuni funzionali, interessanti su questo spazio, siano continui. Per esempio, una successione di curve lisce $y = f_n(x)$ su $[0,1]$ converge a $y = f(x)$ se $\sup |f(x) - f_n(x)| \to 0$ per $n \to \infty$. Ma il funzionale «lunghezza della curva» in questa topologia non è continuo: ad esempio $f_n(x) = (1/n) \sin nx$ tende a $f \equiv 0$, mentre la lunghezza della curva f_n non tende a 1. Per rendere continuo questo funzionale occorre rafforzare la topologia richiedendo, ad esempio, la convergenza uniforme non solo delle funzioni, ma anche delle loro derivate.

In altri casi può risultare utile indebolire la topologia, richiedendo solo la convergenza «in media», ad esempio $\int_0^1 |f(x) - f_n(x)|^2 \, dx \to 0$: qui sono ammesse deviazioni locali arbitrariamente grandi di f_n da f, purché siano concentrate su insiemi di misura piccola. Le «norme» e le «seminorme» mediante le quali si assegnano le topologie sugli SVT generali (cfr. oltre) sono in genere scelte in modo tale da garantire la continuità dei funzionali per noi importanti.

Può accadere che uno SVT, ad esempio uno spazio di funzioni, dopo essere stato costruito risulti non-completo, come non è completo lo spazio dei numeri razionali nella topologia archimedea; allora occorre completarlo. Si dànno delle definizioni precise che è comodo formulare in termini di reti convergenti, le quali generalizzano le successioni convergenti dell'analisi classica.

Sia A un insieme diretto (o filtrante a destra), cioè un insieme su cui sia data una relazione d'ordine $\alpha > \beta$ soddisfacente l'ulteriore condizione: per ogni $\alpha, \beta \in A$ esiste $\gamma \in A$ tale che $\gamma > \alpha$ e $\gamma > \beta$. Ogni applicazione di A in X si dice una (A-)rete su X. È comodo indicare una A-rete con $\{x_\alpha \mid \alpha \in A, x_\alpha \in X\}$, o semplicemente $\{x_\alpha\}$. Se X è uno spazio topologico, $\{x_\alpha\}$ una rete su X, $x \in X$, scriveremo $x_\alpha \to x$ o $\lim x_\alpha = x$, se per ogni intorno $U \ni x$ esiste un $\alpha \in A$ tale che per ogni

$\beta \in A$, $\beta > \alpha$, risulti $x_\beta \in U$. In questo caso si dice che la rete x_α converge al punto x. In uno spazio di Hausdorff una rete può convergere a un solo punto. L'assegnazione di una topologia in X può, a volte, essere sostituita dalla descrizione di tutte le reti, convergenti ai punti di X, poiché un'applicazione $f: X \to Y$ è continua se e solo se da $x_\alpha \to x$ segue che $f(x_\alpha) \to f(x)$ per ogni rete su X.

Una rete $\{x_\alpha\}$ in uno SVT L si dice fondamentale se per ogni intorno dello zero $U \subseteq L$ esiste un α tale che da $\beta > \alpha$ e $\gamma > \alpha$ segue $x_\beta - x_\gamma \in U$. Uno spazio L si dice completo se ogni rete fondamentale ha un limite in esso. La costruzione classica di $\mathbf{R}$ a partire da $\mathbf{Q}$ si trasporta agli SVT e permette, per ogni SVT L, di costruire il suo completamento $\bar{L}$ insieme con un'applicazione continua $L \to \bar{L}$.

Infine, dagli SVT iniziali, del tipo spazi di funzioni, si possono formare nuovi SVT, come gli spazi dei funzionali lineari continui su essi, dei prodotti tensoriali, ecc. Gli effetti della non-unicità della topologia continuano ad agire anche in questi casi; si tornerà in seguito su queste costruzioni.

Norme e seminorme.

Si dice seminorma p su uno spazio vettoriale L su $K = \mathbf{R}$ o $\mathbf{C}$, l'applicazione $p: L \to \mathbf{R}$ che soddisfa le condizioni: 1) $p(x) \geq 0$; 2) $p(x+y) \leq p(x) + p(y)$; 3) $p(\lambda x) = |\lambda| p(x)$ per ogni $x, y \in L$, $\lambda \in K$. La seminorma p si dice norma se, inoltre, da $p(x) = 0$ segue che $x = 0$. Se p è una norma, invece di $p(x)$ spesso si scrive $\|x\|$.

Ogni famiglia di seminorme (p_i) su L definisce una struttura di SVT nella quale una rete $x_\alpha \in L$ converge a $x \in L$ se e solo se $p_i(x_\alpha - x) \to 0$ per ogni i. Se la topologia su L può essere assegnata da una famiglia consistente di una sola norma, L si dice normato.

Uno SVT normato completo si dice spazio di Banach.

Uno spazio di Banach la cui norma $\| \ \|$ soddisfa l'identità $\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$ si chiama spazio di Hilbert.

Esempio. Sia X uno spazio vettoriale, $C(X)$ lo spazio delle funzioni continue su X. Si prenda in considerazione il caso $X = [0,1]$ e per ogni $p > 0$ si ponga per $f \in C[0,1]$

$$\|f\|_p = [\textstyle\int_0^1 |f(x)|^p \, dx]^{1/p}$$
$$\|f\|_\infty = \sup |f|.$$

Si può dimostrare che per ogni $p \geq 1$ o $p = \infty$, $\| \ \|_p$ è una norma. Sia τ_p la topologia definita dalla norma $\| \ \|_p$, $L_p[0,1]$ il completamento di $C[0,1]$ in questa topologia. La norma $\| \ \|_p$ si prolunga su $L_p[0,1]$ e lo rende uno spazio di Banach. Se $p = \infty$ esso coincide con $C[0,1]$; se $p < \infty$ non coincide. Se $p = 2$, $L_2[0,1]$ è uno spazio di Hilbert.

Esempio. Sia $\Omega \subset \mathbf{R}^n$ un aperto, $C^\infty(\Omega)$ lo spazio delle funzioni infinitamente differenziabili su Ω. La topologia abituale su $C^\infty(\Omega)$ è definita dalla convergenza uniforme, con tutte le derivate, su ogni compatto in Ω. Rappresentando Ω come unione $\bigcup_{m=1}^{\infty} K_m$ di sottoinsiemi chiusi e limitati (e perciò compatti), si può assegnare questa topologia mediante il sistema di seminorme

$$p_m(u) = \sup_{x \in K_m} \sum_{|q|=0}^{m} |u^{(q)}(x)|.$$

Qui $x = (x_1, \ldots, x_n)$, $q = (q_1, \ldots, q_n)$, $|q| = q_1 + \ldots + q_n$, $u^{(q)}(x) = \dfrac{\partial^{|q|} u}{\partial x_1^{q_1} \ldots \partial x_n^{q_n}}(x)$.

Lo spazio delle funzioni fondamentali $\mathcal{D}(\Omega) \subset C^\infty(\Omega)$ per definizione è formato dalle funzioni a supporto compatto: ognuna di esse si annulla all'esterno di un opportuno sottoinsieme chiuso e limitato di Ω. Il piú piccolo sottoinsieme chiuso all'esterno del quale la funzione u sia uguale a zero si dice supporto di u, e si indica con $\operatorname{supp} u$.

Se $K \subset \Omega$ è chiuso e limitato, lo spazio $\mathcal{D}_K$ delle funzioni $u \in C^\infty(\Omega)$ con $\operatorname{supp} u \subset K$ è dotato di una topologia per mezzo del sistema di seminorme

$$p_{K,m}(u) = \sup_{x \in K} \sum_{|q|=0}^{m} |u^{(q)}(x)|.$$

Se $K_1 \subset K_2$, si ha $\mathcal{D}_{K_1} \subset \mathcal{D}_{K_2}$, e le topologie su $\mathcal{D}_{K_1}$ e $\mathcal{D}_{K_2}$ sono compatibili. Nello spazio $\mathcal{D}(\Omega) = \bigcup_{K \subset \Omega} \mathcal{D}_K$ si introduce la topologia piú debole che induca su ogni $\mathcal{D}_K$ la topologia descritta sopra. Lo spazio ottenuto si dice spazio delle funzioni fondamentali su Ω. Le funzioni generalizzate su Ω saranno in seguito definite come funzionali lineari continui su $\mathcal{D}(\Omega)$.

È necessario osservare che $\mathcal{D}(\Omega)$ è solo uno tra i possibili spazi di funzioni fondamentali rispetto ai quali, per dualità, si definiscono spazi utili di funzioni generalizzate. Un'altra classe di spazi è descritta nel seguente esempio:

Esempio. Lo spazio $S_\alpha^\beta (\alpha \geq 0, \beta \geq 0)$ per definizione è formato da quelle funzioni infinitamente differenziabili $\varphi(x)$ su $\mathbf{R}$ per cui esistono delle costanti A, B, C (dipendenti da φ) per cui: $|x^k \varphi^{(q)}(x)| \leq C A^k B^q k^{k\alpha} q^{q\beta}$ per ogni k, q.

La topologia su S_α^β s'introduce nel modo seguente. Sia $S_{\alpha,A}^{\beta,B}$ lo spazio $\{\varphi(x) \mid |x^k \varphi^{(q)}(x)| \leq C \bar{A}^k \bar{B}^q k^{k\alpha} q^{q\beta}\}$ per arbitrari $k, q, \bar{A} > A$ e $\bar{B} > B$. S'introduca una topologia su $S_{\alpha,A}^{\beta,B}$ mediante il sistema di norme

$$\|\varphi\|_{m,n} = \sup_{x,k,q} |x^k \varphi^{(q)}(x)| \left[\left(A + \frac{1}{m}\right)^k \left(B + \frac{1}{n}\right)^q k^{k\alpha} q^{q\beta} \right]^{-1}.$$

Esso è completo in questa topologia. Su $S_\alpha^\beta = \bigcup_{A,B} S_{\alpha,A}^{\beta,B}$ si introduce la topologia dell'unione, nella quale una successione di funzioni converge a zero se essa è tutta contenuta in uno degli spazi $S_{\alpha,A}^{\beta,B}$ e converge a zero in questo spazio.

Le costanti α e β impongono delle restrizioni sul modo in cui la funzione e le sue derivate decrescono all'infinito. Se queste restrizioni sono troppo forti, può accadere che lo spazio S_α^β si riduca al solo zero. Un risultato non-banale è che $S_\alpha^\beta \neq \{0\}$ solo per $\alpha + \beta \geq 1$, $\alpha > 0$, $\beta > 0$, e anche per $\beta > 1$, $\alpha = 0$ e $\alpha > 1$, $\beta = 0$. La trasformata di Fourier trasforma S_α^β in S_β^α, essenzialmente perché scambia di posto gli operatori di moltiplicazione per x e quelli di derivazione rispetto a x.

Topologia e convessità.

Sia L uno spazio vettoriale su $\mathbf{R}$ o su $\mathbf{C}$, e $x,y\in L$. L'intervallo di estremi x, y in L è il sottoinsieme $\{\alpha x+(1-\alpha)y\}$, $0\leq\alpha\leq 1$. Un sottoinsieme $U\subset L$ si dice convesso se contiene, insieme con due qualsiasi punti $x,y\in U$, anche l'intero intervallo di estremi x, y. Un sottoinsieme $U\subset L$ si dice equilibrato se da $x\in U$, e $|\lambda|\leq 1$ segue che $\lambda x\in U$. Un sottoinsieme $U\subset L$ si dice assorbente se per ogni $x\in L$ esiste $\lambda>0$ con $\lambda x\in U$.

La connessione tra questi concetti e i precedenti è la seguente: se nello SVT L la topologia è data mediante una famiglia di seminorme (p_i), allora si ha un sistema fondamentale di intorni dello zero in L: $\bigcap_{K=1}^{n}\{x\,|\,p_{i_K}(x)<\varepsilon\}$, che risultano insiemi convessi, equilibrati ed assorbenti. È vero anche il viceversa: se lo SVT L è localmente convesso, cioè ammette una serie di intorni dello zero formata da insiemi convessi, allora la sua topologia può essere assegnata mediante un sistema di seminorme.

La classe degli spazi localmente convessi (SLC) L sarà fondamentale nelle nostre considerazioni. Il suo ruolo nella teoria della dualità si spiega già per il fatto che per ogni SLC L ed ogni punto $x\in L$ esiste un funzionale lineare continuo $f: L\to\mathbf{R}$ tale che $f(x)\neq 0$.

Spazio topologico duale.

Definizione di L'.

Sia L uno SVT su un campo K. Lo spazio topologico L', duale di L, è formato da tutti i funzionali lineari continui $L\to K$. Esso è un sottospazio dello spazio duale algebrico L^*, nella cui definizione non si tien conto della topologia di L.

Topologie in L'.

In L' è naturale considerare almeno due diverse topologie: la debole e la forte. Nella topologia debole una rete $\{f_\alpha\}$ in L' converge ad f se e solo se per ogni $x\in L$ si ha $f_\alpha(x)\to f(x)$ (convergenza puntuale). Nella topologia forte la convergenza è la convergenza uniforme sugli insiemi limitati di L. (Un sottoinsieme $F\subset L$ si dice limitato se per ogni intorno U dello zero esiste $\lambda\in\mathbf{R}$ tale che $F\subset\lambda U$; negli spazi localmente convessi ciò è equivalente alla limitatezza di $f(F)$ per ogni $f\in L'$).

Se L è uno spazio di Banach con norma p, su L' si ha una struttura naturale di spazio di Banach con norma p', per la quale $p'(f)=\sup_{p(x)\leq 1}|f(x)|$. La topologia corrispondente a p' coincide con la topologia forte.

Esempio: distribuzioni. Si dice distribuzione ogni elemento dello spazio $\mathcal{D}(\Omega)'$, dove lo spazio delle funzioni fondamentali $\mathcal{D}(\Omega)$ è stato definito nell'esempio di p. 164. In altre parole, una distribuzione T è un'applicazione lineare $T:\mathcal{D}(\Omega)\to\mathbf{R}$ (o $\mathbf{C}$) tale che per ogni compatto $K\subset\Omega$ esistono $C>0$ e $n\geq 0$ tali che per ogni $u\in\mathcal{D}(\Omega)$ si ha $|T(u)|\leq C\sup_{x\in K}\sum_{|q|=0}^{n}|u^{(q)}(x)|$.

Ad esempio, se la funzione f in Ω è sommabile in ogni compatto, allora ad essa corrisponde la distribuzione $T_f: u\mapsto\int_\Omega u(x)f(x)\,dx$. Tali T_f si dicono regolari. L'applicazione $f\mapsto T_f$, limitata su $\mathcal{D}(\Omega)$, definisce un'immersione canonica $\mathcal{D}(\Omega)\to\mathcal{D}(\Omega)'$, cosicché lo spazio $\mathcal{D}(\Omega)'$ si comporta come un ampliamento dello spazio di funzioni $\mathcal{D}(\Omega)$, il che spiega l'origine del termine russo 'funzione generalizzata'. Si userà questo termine come sinonimo di «distribuzione».

La convergenza delle distribuzioni è la convergenza nella topologia debole: $T_\alpha\to T$ se per ogni $u\in\mathcal{D}(\Omega)$, $T_\alpha(u)\to T(u)$.

Per le distribuzioni regolari T_f con $f\in\mathcal{D}(\Omega)$ è verificata l'identità $T_{\partial f/\partial x_i}(u)=-T_f(\partial u/\partial x_i)$ (integrazione per parti). È naturale esprimere il funzionale a sinistra dell'uguaglianza nella forma $\frac{\partial}{\partial x_i}T_f$, e quello a destra con $-T_f\circ\frac{\partial}{\partial x_i}$. Si osservi che quest'ultimo funzionale è definito per tutti i $T\in\mathcal{D}(\Omega)'$. Ciò conduce alla definizione dell'operatore di derivazione parziale sulle funzioni generalizzate:

$$\frac{\partial^{|q|}}{\partial x_1^{q_1}\ldots\partial x_n^{q_n}}:\mathcal{D}(\Omega)'\to\mathcal{D}(\Omega)':\frac{\partial^{|q|}T}{\partial x^q}(u)=(-1)^{|q|}T\left(\frac{\partial^{|q|}u}{\partial x^q}\right).$$

Tutti questi operatori sono continui nella topologia di $\mathcal{D}(\Omega)'$; i risultati dell'applicazione di questi a T_f, per una funzione f sommabile, ma non necessariamente differenziabile, sono detti derivate generalizzate di f. Questo è uno degli effetti della dualizzazione: la derivazione diventa possibile per una classe molto piú ampia di funzioni, a prezzo del fatto che il risultato della derivazione è un oggetto di natura differente.

Per mezzo delle derivate generalizzate si può descrivere completamente la classe delle funzioni generalizzate finite T, che si definisce nel modo seguente. Si dice supporto di T, supp T, il piú piccolo insieme chiuso in Ω tale che, per $u\in\mathcal{D}(\Omega)$ con supporto al di fuori di esso, sia $T(u)=0$. Una funzione generalizzata T è finita se il suo supporto è compatto in Ω. Ogni funzione generalizzata finita può essere rappresentata come somma finita di derivate generalizzate di funzioni continue con supporto compatto in Ω. Per esempio, per $\Omega=\mathbf{R}$, la funzione delta δ_0, in un intorno dell'origine, coincide con la derivata seconda della funzione che è uguale a zero per $x\leq 0$, e uguale a x su un qualsiasi intervallo $[0,\varepsilon]$, $\varepsilon>0$.

Esempio. Lo spazio $L_p[0,1]'$, nella topologia forte, è naturalmente isomorfo a $L_q[0,1]$, dove q è definito dalla condizione $(1/p)+(1/q)=1$. La forma bilineare $\langle f,g\rangle\mapsto\int_0^1 fg\,dx$, dove $f\in L_p$, $g\in L_q$, definisce contemporaneamente f come funzionale lineare continuo su L_q, e g su L_p; essa definisce anche la corrispondente dualità. Inoltre la norma di Banach del funzionale $g\mapsto\int_0^1 fg\,dx$ coincide con la norma $\|f\|_p$ della funzione f in L_p.

In particolare lo spazio di Hilbert $L_2[0,1]$ è isometrico al suo duale: su $\mathbf{R}$ l'isometria è realizzata dall'applicazione $f\mapsto$ funzionale $(g\mapsto\int_0^1 fg\,dx)$ (su $\mathbf{C}$ occorre prendere $\int_0^1 f\bar{g}\,dx$). Piú in generale, se L è un qualsiasi spazio di Hilbert su $\mathbf{C}$ con norma $\|\ \|$, su di esso si ha una forma $\mathbf{R}$-bilineare canonica: il prodotto scalare $(x,y)=(1/4)[\|x+y\|^2-\|x-y\|^2+i\|x-iy\|^2-i\|x+iy\|^2]$. Rispetto ad x

essa è lineare anche su **C**, e rispetto ad y è antilineare su **C**: $(x, ay) = \bar{a}(x, y)$. Inoltre essa è continua e non-degenere in entrambi gli argomenti: invero $(x, x) = \|x\|^2$. Perciò essa definisce un'applicazione canonica $L \to L' : x \mapsto \{y \mapsto (y, x)\}$, che è un isomorfismo (antilineare in **C**).

Nel seguito occorreranno concetti fondamentali della teoria della misura.

Misura e integrale. Sia X un insieme. Si dice σ-algebra B su X una famiglia di sottoinsiemi di X, chiusa rispetto ad unioni numerabili, intersezioni numerabili e complementi. La piú piccola σ-algebra, contenente tutti gli insiemi aperti di uno spazio topologico X, si chiama σ-algebra di Borel.

Si dice **misura** μ sulla σ-algebra degli insiemi μ-misurabili B (o, per brevità, su X) un'applicazione $\mu : B \to \mathbf{R}$ (o **C**) tale che da $E_1 \cap E_2 = \varnothing$ segue che $\mu(E_1 \cup E_2) = \mu(E_1) + \mu(E_2)$. La misura μ si dice numerabilmente additiva se da $E_i \cap E_j = \varnothing$ per $i \neq j$ segue che $\mu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)$, dove la serie a destra converge assolutamente. La misura μ è positiva se $\mu(E) \geq 0$ per ogni $E \in B$. Si dice variazione della misura numerabilmente additiva μ la misura $|\mu|(E) = \sup \sum |\mu(E_i)|$, dove il sup è preso su tutte le partizioni $E = \bigcup_{K=1}^{\infty} E_K$, $E_K \in B$, $E_i \cap E_j = \varnothing$ per $i \neq j$. La misura μ è detta completa se da $E_1, E_2 \in B$, $E_1 \subset E \subset E_2$, $|\mu|(E_2 \setminus E_1) = 0$ segue che $E \in B$. Ogni misura numerabilmente additiva si estende a una misura completa (mediante un opportuno ampliamento dell'algebra B).

La consueta misura su **R** (che, sugli intervalli, coincide con la lunghezza), non soddisfa alla condizione che la serie $\sum_{i=1}^{\infty} \mu(E_i)$ converga (necessariamente) per $E_i \cap E_j = \varnothing$. Ma **R** può essere rappresentato come unione di una famiglia numerabile di semiintervalli disgiunti, su ciascuno dei quali la consueta misura di Borel soddisfa le nostre definizioni. Nel caso generale, se $X = \bigcup_{i=1}^{\infty} X_i$ (unione disgiunta), e su ogni X_i è data una misura μ_i, allora la misura $\mu = \sum_{i=1}^{\infty} \mu_i$ si dice σ-finita. Essa è definita per quegli $E \subset X$ per i quali $E \cap X_i$ è μ_i-misurabile, e la serie $\sum_i \mu(E \cap X_i)$ converge assolutamente.

Una funzione $f : X \to \mathbf{R}$ (o **C**) si dice μ-misurabile se la f-controimmagine di un qualsiasi insieme aperto è misurabile. Una funzione misurabile si dice semplice se assume un insieme al piú numerabile di valori distinti $\{f_K\}$. Se la somma $\sum_K f_K \mu(x \mid f(x) = f_K)$ converge assolutamente, essa si chiama integrale della funzione semplice f rispetto alla misura μ, e si indica con $\int_X f(x)\, d\mu(x)$ o $\int f\, d\mu$. Una funzione misurabile $f \geq 0$ si dice μ-integrabile se esiste $C > 0$ tale che per tutte le funzioni semplici $g \leq f$ si ha $|\int_X g\, d\mu| < C$. In questo caso si dice integrale $\int f\, d\mu$ l'estremo superiore degli integrali $\int_X g\, d\mu$ rispetto a tutte le funzioni semplici $g \leq f$.

Una funzione misurabile complessa f si dice μ-integrabile o sommabile se $|f|$ è $|\mu|$-integrabile. L'integrale $\int_X f\, d\mu$ è univocamente definito su tali funzioni dalle condizioni di linearità rispetto ad f e μ.

Con questi dati è possibile continuare la descrizione degli spazi topologici duali.

Esempio: le misure come funzionali continui. Sia X uno spazio di Hausdorff localmente compatto, $K(X)$ lo spazio delle funzioni continue su X a supporto compatto. La topologia su $K(X)$ è data dalla definizione di convergenza: $f_\alpha \to f$ se esiste un compatto $K \subset X$ tale che $K \supseteq \operatorname{supp} f_\alpha, \operatorname{supp} f$ e f_α converge uniformemente ad f.

D'altra parte si consideri lo spazio delle misure regolari su X, cioè delle misure che godono della proprietà: per ogni insieme misurabile E con $\mu(E) < \infty$ e ogni $\varepsilon > 0$ esistono un compatto K e un insieme aperto U tali che $K \subset E \subset U$ e $|\mu|(U \setminus K) < \varepsilon$.

Ogni misura regolare μ definisce su $K(X)$ un funzionale lineare continuo $f \mapsto \int_X f\, d\mu$, e tutti gli elementi dello spazio duale $K(X)'$ si ottengono in tale modo.

Nello spazio delle distribuzioni $\mathfrak{D}(\Omega)'$ si trovano i funzionali $T_\mu : u \mapsto \int_\Omega u(x)\, d\mu(x)$, corrispondenti alle misure σ-finite μ su Ω. La funzione delta di Dirac, $\delta_a (a \in \Omega)$, per esempio, corrisponde alla misura unitaria concentrata nel punto a: $\mu(E) = 1$ se $a \in E$, $\mu(E) = 0$ altrimenti.

Esempio: gli spazi $L_p(X, \mu)$. Se X è uno spazio con misura μ, è possibile definire $L_p(X, \mu)$, per $p \geq 1$, come il completamento dello spazio delle funzioni su X con norma finita $\|f\|_p = |\int_X |f(x)|^p\, d|\mu|(x)|^{1/p}$. Tutti i risultati formulati sopra per gli spazi $L_p[0, 1]$ restano veri in questo contesto piú generale. In particolare $L_p(X, \mu)'$ è isomorfo a $L_q(X, \mu)$, dove $(1/p) + (1/q) = 1$.

Riflessività.

Se L è uno spazio lineare normato, allora L' è uno spazio di Banach, e si può costruire lo spazio L'', biduale di L. Si ha, come nel caso a dimensione finita, un'applicazione lineare canonica $\beta : L \to L''$: $\beta(x)$ è $f(x)$ come funzione di $f \in L'$. Si può dimostrare che l'applicazione β è una isometria di L con $\beta(L)$, e $\beta(L)$ è denso in L'' nella topologia debole, ma non coincide necessariamente con L''.

Lo spazio L si dice riflessivo se β è una isometria di L con L''. Da quanto osservato a p. 167 si deduce che gli spazi di Hilbert sono riflessivi. Analogamente sono riflessivi gli spazi $L_p(X)$ per $p > 1$.

Uno dei criteri di riflessività di L consiste nella condizione che ogni funzionale $f \in L'$ raggiunga il suo estremo superiore sulla sfera unitaria $\|x\| = 1$ in L. Un'altra condizione è la compattezza della sfera unitaria $\|x\| \leq 1$ nella cosiddetta topologia debole in L. In questa topologia una base di intorni dello zero è formata da tutti i possibili insiemi del tipo $\{x \in L \mid |f_1(x)|, \ldots, |f_n(x)| < \varepsilon\}$, dove $\varepsilon > 0$ e $f_1, \ldots, f_n \in L'$ sono funzionali arbitrari.

Operatori lineari, prodotti tensoriali e nuclearità. Se L_1, L_2 sono spazi vettoriali a dimensione finita, si ha l'isomorfismo canonico $\operatorname{Hom}(L_1, L_2) \simeq L_1^* \otimes L_2$. In esso sono unificate due idee di dualità: dualità degli spazi vettoriali e dualità secondo Kan dei due funtori Hom e $\otimes$ (cfr. §§ 1.2 e 2.2). Il trasporto di questa identità al caso di infinite dimensioni non è banale, dato che la definizione dell'analogo topologico del prodotto tensoriale non è univoca, e l'applicazione

di questo nello spazio degli operatori cessa di essere un isomorfismo. Lo studio degli oggetti che si presentano naturalmente in questo contesto conduce all'individuazione dell'importante classe degli operatori e degli spazi nucleari.

Nel seguito si considereranno successivamente le varianti topologiche degli spazi $\mathrm{Hom}(L_1, L_2)$, $L_1^* \otimes L_2$ e le connessioni tra loro.

Spazio degli operatori.

Siano L_1, L_2 due SVT. In questo caso, invece di $\mathrm{Hom}(L_1, L_2)$, si considerano le applicazioni lineari continue $L_1 \to L_2$ (operatori lineari da L_1 in L_2). Gli operatori lineari formano uno spazio vettoriale, che si indica con $\mathfrak{L}(L_1, L_2)$. Se L_1 ed L_2 sono normati, anche in $\mathfrak{L}(L_1, L_2)$ si può definire una norma: $\|A\| = \sup_{\|x\|=1} \|A(x)\|$. Nel caso generale, su $\mathfrak{L}(L_1, L_2)$ si hanno alcune topologie naturali, sulla cui definizione non ci si soffermerà.

Lo spazio topologico duale L' è, evidentemente, $\mathfrak{L}(L, K)$.

Operatore aggiunto.

Sia $A \in \mathfrak{L}(L_1, L_2)$. L'operatore $A^* : L_2^* \to L_1^*$, aggiunto algebrico di A, può essere ristretto al sottospazio topologico duale $L_2' \subset L_2^*$. Si ottiene un operatore $A' : L_2' \to L_1'$ che si dice aggiunto topologico di A.

Prodotti tensoriali di spazi hilbertiani e normati.

Siano (L_1, p_1), (L_2, p_2) spazi normati. Se questi sono hilbertiani, sul prodotto tensoriale algebrico $L_1 \otimes L_2$ è definita un'unica norma hilbertiana che soddisfa la condizione $p(x_1 \otimes x_2) = p_1(x_1) p_2(x_2)$ per ogni $x_1 \in L_1$, $x_2 \in L_2$. Rispetto a questa norma lo spazio $L_1 \otimes L_2$ è separato, ma non completo. Il suo completamento $(L_1 \otimes L_2)^-$ si dice prodotto tensoriale hilbertiano degli spazi L_1 e L_2.

Nel caso generale la condizione $p(x_1 \otimes x_2) = p_1(x_1) p_2(x_2)$ non definisce univocamente una norma su $L_1 \otimes L_2$. Tutte le norme con questa proprietà si chiamano norme tensoriali. Tra queste ne esiste una massima, $p_1 \hat{\otimes} p_2$, e una minima $p_1 \underset{\otimes}{\otimes} p_2$. Cioè $p_1 \hat{\otimes} p_2 = \inf \sum_i p_1(x_i) p_2(y_i)$, dove inf è preso rispetto a tutte le rappresentazioni di z nella forma $z = \sum x_i \otimes y_i$; analogamente $(p_1 \underset{\otimes}{\otimes} p_2) = \sup |f(z)|$, dove sup è preso su tutte le $f = f_1 \otimes f_2$, $f_1 \in L_1'$, $f_2 \in L_2'$, $p_i'(f_i) \leq 1$. I corrispondenti completamenti si indicano con $L_1 \hat{\otimes} L_2$ e $L_1 \underset{\otimes}{\otimes} L_2$. Si ha un'immersione naturale $L_1 \hat{\otimes} L_2 \subset L_1 \underset{\otimes}{\otimes} L_2$.

Connessione tra prodotti tensoriali ed operatori.

Sia definita l'applicazione canonica $L_1' \otimes L_2 \to \mathfrak{L}(L_1, L_2)$ che all'elemento $f \otimes x$ ($f \in L_1'$, $x \in L_2$) associa l'operatore di rango 1, $y \mapsto f(y) x$ (si dice rango di un operatore lineare la dimensione della sua immagine). L'immagine di $L_1' \otimes L_2$ consiste di tutti gli operatori di rango finito. L'applicazione canonica descritta si estende ad un'applicazione $L_1' \otimes L_2 \to \mathfrak{L}(L_1, L_2)$ che, in generale, non è né iniettiva né suriettiva. La sua immagine coincide con la chiusura in norma dello spazio degli operatori di rango finito. Tutti gli operatori di questo tipo sono completamente continui, cioè trasformano insiemi limitati in insiemi con chiusura compatta.

L'immagine del prodotto tensoriale hilbertiano $(L' \otimes L)^-$ in $\mathfrak{L}(L, L)$ consiste dei cosiddetti operatori di Hilbert-Schmidt.

Nell'immersione naturale $L_1' \hat{\otimes} L_2 \subset L_1' \underset{\otimes}{\otimes} L_2$ gli elementi di $L_1' \hat{\otimes} L_2$ sono trasformati in elementi dello spazio di operatori $\mathfrak{L}(L_1, L_2)$, che si dicono nucleari. Essi hanno la forma $x \mapsto \sum_{n=1}^{\infty} f_n(x) x_n$, dove $f_n \in L_1'$, $x_n \in L_2$ e $\sum_{n=1}^{\infty} \|f_n\| \, \|x_n\| < \infty$.

Se L è uno spazio di Hilbert, si ha il seguente criterio di nuclearità di un operatore $A : L \to L$. Sia (x_α) una base ortonormale di L, cioè tale che $\|x_\alpha\| = 1$ e $(x_\alpha, x_\beta) = 0$ per ogni $\alpha \neq \beta$ e $(y, x_\alpha) = 0$ per ogni α implica che $y = 0$. La nuclearità di A equivale alla convergenza della serie $\sum_\alpha (A x_\alpha, x_\beta)$ per ogni base ortonormale. La somma di questa serie non dipende dalla scelta della base, e si dice traccia dell'operatore A.

Spazi nucleari.

Si supponga ora che la topologia su L_1 sia data dal sistema di seminorme $\{p_\alpha\}$, e in L_2 dal sistema di seminorme $\{q_\beta\}$. Le seminorme $p_\alpha \hat{\otimes} q_\beta$ e $p_\alpha \underset{\otimes}{\otimes} q_\beta$ in $L_1 \otimes L_2$ sono definite esattamente dalle stesse relazioni di prima. I completamenti di $L_1 \otimes L_2$ rispetto ai sistemi di seminorme $\{p_\alpha \hat{\otimes} q_\beta\}$ e $\{p_\alpha \underset{\otimes}{\otimes} q_\beta\}$ saranno indicati rispettivamente con $L_1 \hat{\otimes} L_2$ e $L_1 \underset{\otimes}{\otimes} L_2$. Come sopra, si ha un'immersione naturale $L_1 \hat{\otimes} L_2 \subset L_1 \underset{\otimes}{\otimes} L_2$.

Uno SLC L si dice nucleare se per ogni SLC M risulta $L \hat{\otimes} M = L \underset{\otimes}{\otimes} M$.

Tra gli spazi normati solo quelli a dimensione finita sono nucleari. Se sullo SLC L è definito un sistema numerabile di seminorme (p_k), allora la condizione di nuclearità per L è la seguente: per ogni k esiste un $j > k$ tale che l'applicazione di identificazione $(L, p_j) \to (L, p_k)$ sia nucleare.

Gli spazi nucleari posseggono una serie di buone proprietà topologiche. In essi la topologia debole e la forte coincidono, gli insiemi chiusi e limitati sono compatti, come nel caso a dimensione finita. La categoria degli spazi nucleari è chiusa rispetto ai limiti proiettivi ed induttivi, e al passaggio a sottospazi chiusi e a spazi quoziente rispetto a questi. Infine, ogni operatore lineare continuo da uno spazio nucleare o in uno spazio nucleare è nucleare. Piú precisamente, l'applicazione canonica $L_1' \hat{\otimes} L_2 = L_1' \underset{\otimes}{\otimes} L_2 \to \mathfrak{L}(L_1, L_2)$ è un isomorfismo, cosicché si ottiene un'analogia completa con il caso a dimensione finita. L'importanza del concetto di nuclearità non è determinata solo da queste proprietà. L'individuazione di questo concetto ha rappresentato il compimento di un certo stadio della geometrizzazione dell'analisi funzionale lineare.

L'introduzione dei concetti di spazio di Hilbert inizialmente, e poi di spazio di Banach, sono state le prime pietre miliari di questo stadio. Ma il contesto degli spazi di Banach è risultato ristretto per esempi importanti e naturali quali gli spazi delle funzioni fondamentali e delle distribuzioni. L'assiomatizzazione delle loro proprietà topologiche ha condotto alla definizione degli spazi localmente convessi, ma questa categoria è risultata troppo ampia per tutta una serie di scopi. Ora si ritiene che gli spazi nucleari, di cui fanno parte tutti gli spazi importanti di funzioni fondamentali e generalizzate, costituiscano la classe piú na-

turale rispetto a tutta una serie di proprietà. La complessità della loro definizione è bilanciata dalla semplicità dell'«algebra multilineare» in questa classe. Un ulteriore esempio delle buone proprietà degli spazi nucleari è l'esistenza di una «analisi di Fourier» su di essi. Ogni rappresentazione unitaria di $\mathbf{R}^n$ può essere rappresentata come integrale di Fourier rispetto ad una certa misura sullo spazio duale $\mathbf{R}^{*n}$. Se, invece di $\mathbf{R}^n$, si considera uno SLC, questo risultato rimane vero se e solo se tale spazio è nucleare.

3.3. Algebre di Banach e loro spettri.

La geometria si occupa degli spazi, l'analisi delle funzioni su di essi. La relazione spazio/funzioni ha le caratteristiche della dualità già per il fatto che «inverte le frecce»: un'applicazione di spazi corrisponde al trasporto delle funzioni nella direzione opposta (cfr. § 2). Punto di partenza possono essere tanto gli spazi quanto le funzioni, eventualmente in senso generalizzato. La teoria degli schemi in geometria algebrica ha origine dalla formulazione precisa dell'idea che ogni anello commutativo è un anello di funzioni su un appropriato oggetto geometrico: il suo spettro. Questa costruzione è stata preceduta dalla fondazione della teoria delle algebre di Banach e dei loro spettri nel contesto dell'analisi. Nella sua accezione generale lo spettro di un anello è quello spazio naturale sul quale gli elementi dell'anello siano funzioni. Si dànno qui di seguito alcune definizioni precise.

Algebre di Banach.

Un'algebra, dotata di unità sul campo dei numeri complessi, si dice di Banach se è dotata di una struttura di spazio di Banach, con norma $\| \ \|$, per la quale $\|xy\| \leq \|x\| \, \|y\|$, $\|1\| = 1$, e la moltiplicazione è continua.

Nel seguito ci si limiterà a considerare algebre di Banach commutative.

Caratteri e spettro.

Sia A un'algebra di Banach commutativa. Si dice carattere di A un omomorfismo continuo non nullo $\chi : A \to \mathbf{C}$. Si dice ideale massimale in A un ideale $\mathfrak{i} \subset A$ tale che tra $\mathfrak{i}$ ed A non vi siano ideali intermedi. A partire da ciascun carattere χ si può costruire un ideale massimale $\mathfrak{i}$, il nucleo di χ; a partire da ogni ideale massimale, un carattere $\chi : A \to A/\mathfrak{i} \simeq \mathbf{C}$. Si dice spettro $M(A)$ dell'algebra A l'insieme dei suoi ideali massimali, o l'insieme dei suoi caratteri: per quanto detto sopra essi sono in corrispondenza biunivoca.

Esempio. L'anello delle funzioni continue $C(X)$ su un compatto X con $\|f\| = \sup_{x \in X} |f(x)|$ è un'algebra di Banach. Ogni punto $x \in X$ definisce *a*) il carattere $\chi_x : f \mapsto f(x)$; *b*) l'ideale massimale $\mathfrak{i}_X = \{f \mid f(x) = 0\}$. In tal modo si ha un'applicazione $X \to M(C(X))$. Si può dimostrare che essa è biiettiva, cioè che X si ricostruisce come insieme a partire dal suo spettro $M(C(X))$. In effetti si può ricostruire anche la topologia.

Topologia sullo spettro.

Ogni carattere di A è un elemento dello spazio di Banach A', duale topologico di A: si ha perciò un'immersione naturale $M(A) \subset A'$. Si doti $M(A)$ della topologia indotta dalla topologia debole su A'. Si può dimostrare che $M(A)$ è chiuso e limitato e perciò compatto. Nell'esempio precedente l'applicazione canonica $X \to M(C(X))$ è un omeomorfismo.

Esempio. $A = \{x = (x_0, x_1, \ldots) \mid \|x\| = \sum_{i=0}^{\infty} |x_i| < \infty\}$ con la moltiplicazione $(xy)_n = \sum_{k=0}^{\infty} x_k y_{n-k}$. Lo spettro $M(A)$ è isomorfo al cerchio unitario complesso $\{|z| \leq 1 \mid z \in \mathbf{C}\}$. Il punto z corrisponde al carattere $x \mapsto \sum_{i=0}^{\infty} x_i z^i$.

Relazioni di dualità.

L'applicazione M (spettro) è un funtore dalla categoria delle algebre di Banach nella categoria duale della categoria degli spazi topologici compatti. È naturale considerare lo spazio $M(A)$ come oggetto duale dell'algebra A.

Per ogni A si ha un'applicazione naturale di algebre di Banach $A \to C(M(A))$: $x \mapsto \hat{x}$, dove $\hat{x}(\chi) = \chi(x)$ per ogni $x \in A$, $\chi \in M(A)$. Il secondo esempio mostra che esso non è necessariamente un isomorfismo: in esso l'immagine di A è formata dalle funzioni analitiche nel cerchio, ma non da tutte le funzioni analitiche. Una dualità precisa si ottiene soltanto per le algebre $C(X)$ con X compatto.

3.4. Gruppi topologici e trasformata di Fourier.

La teoria della dualità per i gruppi topologici ha una forma compiuta solo nel caso commutativo. Per gruppi non-commutativi un'appropriata definizione di oggetto duale non è stata trovata, a quanto pare, neppure nel caso finito. Qui ci si limiterà perciò ad una breve discussione della dualità di Pontrjagin e dello spazio $\hat{G}$ delle rappresentazioni unitarie irriducibili.

Lo spazio $\hat{G}$.

Gruppi topologici.

Sia G un gruppo topologico, cioè un insieme G dotato della struttura di gruppo e di spazio topologico in maniera tale che sia continua l'applicazione $G \times G \to G : \langle x, y \rangle \mapsto xy^{-1}$.

Rappresentazioni.

Si dice rappresentazione T del gruppo topologico G in uno SVT L una rappresentazione algebrica $G \times L \to L : \langle g, x \rangle \mapsto T(g)x$, continua rispetto a $\langle g, x \rangle$. Una rappresentazione T si dice unitaria se L è uno spazio di Hilbert e tutti gli operatori $T(g)$ sono unitari, cioè conservano il prodotto scalare hilbertiano. Due rappresentazioni unitarie T_1, T_2 negli spazi L_1, L_2 sono isomorfe se esiste una isometria di L_1 con L_2 che trasforma $T_1(g)$ in $T_2(g)$ per ogni $g \in G$. Una rappre-

sentazione T si dice topologicamente irriducibile se nel suo spazio L non esiste un sottospazio proprio chiuso invariante rispetto a tutti i $T(g)$.

Lo spazio $\hat{G}$.

Con $\hat{G}$ si indica l'insieme delle classi delle rappresentazioni unitarie irriducibili del gruppo G, con la seguente topologia. Sia T una rappresentazione unitaria di G in L, $[T]$ la sua classe di equivalenza. Una base di intorni del punto $[T]$ in $\hat{G}$ è costituita dagli insiemi indicati con $U(K, \xi_1, \ldots, \xi_n, \varepsilon)$, dove $K \subset G$ è un compatto, $\xi_1, \ldots, \xi_n \in L$ è un qualsiasi insieme finito di vettori, $\varepsilon > 0$ è un numero reale. Una rappresentazione T_1 nello spazio L_1 appartiene a tale insieme se e solo se esistono $\eta_1, \ldots, \eta_n \in L_1$ tali che per ogni $g \in K$, $i, j = 1, \ldots, n$, si ha $|(T(g)\xi_i, \xi_j) - (T_1(g)\eta_i, \eta_j)| < \varepsilon$.

Lo spazio $\hat{G}$ si dice spazio duale del gruppo G. Nel caso in cui quest'ultimo sia commutativo e localmente compatto, lo stesso $\hat{G}$ ha la struttura di gruppo commutativo localmente compatto, e si dice duale di G secondo Pontrjagin (cfr. oltre). Nel caso generale $\hat{G}$ può essere uno spazio topologico con una struttura abbastanza complessa, in particolare non-separato. Osserviamo ancora che la restrizione alle rappresentazioni unitarie non è necessariamente la migliore; per i gruppi di Lie sono utili anche altre varianti.

Esempio. G sia il gruppo diedrale infinito. Ciò significa che G è generato da due generatori a, t con le relazioni $t^2 = 1$, $tat^{-1} = a^{-1}$. In tale caso $\hat{G}$ è omeomorfo all'intervallo $[0, 1]$, con gli «estremi raddoppiati»: $\hat{G} = (0, 1) \cup \{0_1\} \cup \cup \{0_2\} \cup \{1_1\} \cup \{1_2\}$, dove $(0, 1)$ ha la topologia abituale, ma per $t \to 0+$ in $[0, 1]$ si ha $t \to 0_1$ e 0_2 in $\hat{G}$, e per $t \to 1-$ in $[0, 1]$ abbiamo $t \to 1_1$ e 1_2 in $\hat{G}$.

Esempio. $G = SL(2, \mathbf{C})$, gruppo delle matrici 2×2 con elementi in $\mathbf{C}$ e con determinante unitario. Qui $\hat{G}$ è identificato al sottoinsieme di $\mathbf{R}^2$: $\{\langle n, r\rangle \mid n \geq 0$ intero, $r \in R$, $r \geq 0$ per $n = 0\}$, e la topologia su $\hat{G}$ coincide con quella indotta da $\mathbf{R}^2$, con l'eccezione che, quando $t \to \langle -1, 0\rangle$ nel senso consueto, t tende contemporaneamente anche a $\langle 2, 0\rangle$ in $\hat{G}$.

Dualità di Pontrjagin.

Struttura di gruppo su $\hat{G}$.

Sia G un gruppo commutativo localmente compatto. In questo caso $\hat{G}$ è formato esattamente dalle classi di rappresentazioni unitarie unidimensionali: le rappresentazioni di dimensione > 1 sono riducibili. Come nel § 1, ciò permette di identificare $\hat{G}$ con lo spazio dei caratteri unitari di G, cioè gli omomorfismi continui $G \to \mathbf{C}_1 = \{z \in \mathbf{C} \mid |z| = 1\}$. La topologia su $\hat{G}$ è la topologia della convergenza uniforme sui compatti.

La teoria della dualità di Pontrjagin consiste di due parti: una riguarda i gruppi, l'altra le funzioni su di essi (trasformata di Fourier). La prima parte può essere descritta come una serie di affermazioni, esattamente parallele a quelle fatte nel § 1 per i gruppi abeliani finiti. Per non ripetere, si utilizza il linguaggio funtoriale del § 2, nel quale si possono formulare concisamente tutti i fatti principali.

TEOREMA. *a) L'applicazione $X: G \mapsto \hat{G}$ è un funtore dalla categoria dei gruppi localmente compatti nella categoria ad essa duale.*

b) Il funtore X è una equivalenza di categorie, ed il funtore X^0 è il suo inverso, nel senso che XX^0 e X^0X sono equivalenti al funtore identico.

Struttura della categoria dei gruppi abeliani localmente compatti.

La decifrazione delle affermazioni del teorema e la loro dimostrazione contengono come casi particolari tutte le considerazioni formulate nel § 1 per i gruppi abeliani finiti. In particolare l'applicazione canonica $G \to \hat{\hat{G}}: g \mapsto \chi(g)$ come funzione di χ è un isomorfismo topologico.

In effetti si può descrivere in modo molto soddisfacente la struttura della categoria LCAb e del funtore X nel modo seguente.

a) La categoria LCAb contiene i gruppi «elementari» $\mathbf{R}$, $\mathbf{C}_1 \simeq \mathbf{R}/\mathbf{Z}$, $\mathbf{Z}$ e $\mathbf{Z}/n\mathbf{Z}$; inoltre $\hat{\mathbf{R}} \simeq \mathbf{R}$, $\hat{\mathbf{C}}_1 \simeq \mathbf{Z}$, $\hat{\mathbf{Z}} \simeq \mathbf{C}_1$, $(\mathbf{Z}/n\mathbf{Z})\hat{} \simeq \mathbf{Z}/n\mathbf{Z}$.

b) Nella categoria LCAb sono definite le operazioni di somma, di prodotto, di limite proiettivo ed induttivo (con famiglia arbitraria di indici).

c) La categoria LCAb è equivalente alla piú piccola delle sue sottocategorie contenente i gruppi elementari e chiusa rispetto alle operazioni descritte.

d) Il funtore di dualità di Pontrjagin X trasforma gruppi elementari in elementari (a meno di isomorfismi), e prodotti e limiti proiettivi rispettivamente in somme e limiti induttivi, e viceversa. Il gruppo G è compatto se e solo se $\hat{G}$ è discreto.

Trasformata di Fourier.

Integrazione invariante.

Su ogni gruppo localmente compatto G con una base numerabile (di intorni dell'identità) è definita una misura di Borel regolare σ-finita non nulla invariante rispetto alle traslazioni sinistre: $G \to G: x \mapsto gx$, $g \in G$. Essa è definita univocamente a meno di una costante moltiplicativa, ed è detta misura di Haar (sinistra). Un'affermazione analoga è vera per le misure invarianti a destra. La scelta della costante moltiplicativa si dice normalizzazione; se G è compatto, le misure spesso si normalizzano in modo che la misura di G sia uguale a 1. In notazione integrale invece di $d\mu(g)$ per la misura di Haar sinistra scriveremo $d_l g$, per quella destra $d_r g$, e per quella bilaterale (in particolare per i gruppi abeliani) dg.

Esempi di misure invarianti sono: dx su $\mathbf{R}$, $x^{-1}dx$ su $\mathbf{R}^*$; la misura μ su $\mathbf{Z}_p$, per la quale la misura del sottogruppo aperto $U \subset \mathbf{Z}_p$ è uguale a $1/[\mathbf{Z}_p : U]$.

Trasformata di Fourier.

Sia G un gruppo localmente compatto abeliano. La trasformata di Fourier può essere definita come l'applicazione $F: L_1(G, dg) \to L_1(\hat{G}, d\chi): f \mapsto \hat{f}$, dove $\hat{f}(\chi) = \int_G f(g)\chi(g)\,dg$.

Nel novero delle affermazioni sulla dualità rientrano i seguenti fatti:

Le misure di Haar dg e $d\chi$ possono essere normalizzate in modo tale che l'applicazione inversa di F possa essere definita dalla formula $f(g)=\int_{\hat{G}}\hat{f}(\chi)\chi(g)^{-1}\,d\chi$.

Tali misure si dicono duali. Per esempio, se G è compatto e $\int_G dg=1$, allora $\hat{G}$ è discreto e la misura duale di ogni punto di $\hat{G}$ è uguale ad 1.

Nel caso $G=\mathbf{R}/\mathbf{Z}$ si ottiene la costruzione classica della serie di Fourier $(\mathbf{R}/\mathbf{Z})\hat{}\simeq\mathbf{Z}\simeq\{e^{-2\pi inx}\}$ e $\hat{f}$, come funzione su $\mathbf{Z}$, cioè come successione $\{\hat{f}_n\}$, è la successione dei coefficienti di Fourier della funzione f.

Nel caso $G=\mathbf{R}$ si ottiene la costruzione classica dell'integrale di Fourier $\hat{\mathbf{R}}=\mathbf{R}$ (accoppiamento $(p, x)\mapsto e^{-2\pi ipx}$) e $\hat{f}(p)=(1/\sqrt{2\pi})\int_{-\infty}^{+\infty}f(x)e^{-2\pi ipx}$.

Negli stessi casi è verificata l'uguaglianza di Parseval:

$$\int_G |f(g)|^2\,dg=\int_{\hat{G}}|\hat{f}(\chi)|^2\,d\chi.$$

In particolare, per $G=\mathbf{R}/\mathbf{Z}$, si trova l'uguaglianza classica:

$$\int_0^1 |f(x)|^2\,dx=\sum_{n=-\infty}^{\infty}|\hat{f}_n|^2.$$

Da ciò segue che la restrizione della trasformazione di Fourier su $L_2(G, dg)$ si estende ad una isometria degli spazi di Hilbert $L_2(G, dg)\simeq L_2(\hat{G}, d\chi)$.

Esempio: osservabili coniugate in meccanica quantistica e principio di indeterminazione. Si ricordi che un sistema quantistico è descritto dallo spazio di Hilbert dei suoi stati, $\mathcal{H}$, e dagli operatori lineari continui autoaggiunti A agenti su di esso (osservabili). Le osservabili A corrispondono a grandezze fisiche classiche (energia, impulso, coordinate) o non-classiche (spin). Uno stato puro del sistema è un raggio complesso in $\mathcal{H}$; se ψ è un vettore in esso, con $\|\psi\|=1$, la misura della grandezza A nello stato ψ dà, in generale, un risultato indefinito, per il quale si può predire soltanto una distribuzione di probabilità. In particolare il valor medio di una misura è uguale a $(A\psi, \psi)=\langle A\rangle_\psi$, e la dispersione è uguale a $(\Delta A)_\psi=\|[A-\langle A\rangle_\psi]\psi\|$.

Le osservabili $A_1, \ldots, A_n$ formano un sistema completo di osservabili compatibili se $\mathcal{H}$ può avere una realizzazione come $L_2(M, \mu)$, dove $M\subseteq\mathbf{R}^n$, μ è una misura su $\mathbf{R}^n$, e $A_i\psi=$ moltiplicazione per la i-esima coordinata di $\mathbf{R}^n$. Questa realizzazione si dice adattata alla famiglia $A_1, \ldots, A_n$; gli operatori A_i, A_j commutano.

Per esempio, le osservabili di posizione X_1, X_2, X_3 di una particella libera (senza spin e nell'approssimazione non relativistica) agiscono come operatori di moltiplicazione per x_1, x_2, x_3 sulle funzioni lisce a supporto compatto $\psi(x_1, x_2, x_3)$ di $H=L_2(\mathbf{R}^3)$ (convenientemente estesi ad operatori autoaggiunti su tutto H). In questa «rappresentazione coordinata» le osservabili corrispondenti alle proiezioni degli impulsi P_1, P_2, P_3 sono le estensioni degli operatori

$$P_j=-ih\frac{\partial}{\partial x_j}$$

(h è la costante di Planck). Inversamente, nella rappresentazione degli impulsi sulle funzioni $\hat{\psi}(p_1, p_2, p_3)$ le osservabili di posizione hanno la forma

$$\hat{X}_j=ih\frac{\partial}{\partial p_j},$$

e gli impulsi $\hat{P}_j=$ moltiplicazione per p_j. L'isomorfismo delle due rappresentazioni, che identifica le funzioni d'onda ψ di uno stesso stato fisico, è dato dalla trasformata di Fourier dal gruppo localmente compatto $\mathbf{R}^3$ (coordinate) nel suo duale $\hat{\mathbf{R}}^3$ (impulsi):

$$\hat{\psi}(p_1, p_2, p_3)=F\psi(p_1, p_2, p_3)=\frac{1}{(2\pi)^{3/2}}\int_{\mathbf{R}^3}e^{-i(x_1p_1+x_2p_2+x_3p_3)/h}\,\psi(x)\,dx_1\,dx_2\,dx_3.$$

Corrispondentemente si trasformano le osservabili: $\hat{X}_j=FX_jF^{-1}$, $\hat{P}_j=FP_jF^{-1}$.

Ciò spiega uno degli aspetti dell'uso del termine 'osservabili coniugate' in relazione alla coppia coordinate-impulsi: le corrispondenti osservabili sono coniugate rispetto alla trasformazione di Fourier. Per lo stesso motivo si dicono coniugate le rappresentazioni delle coordinate e quelle degli impulsi.

Se A e B sono due osservabili, le dispersioni dei risultati di una loro misura nello stato ψ sono connesse dalla disuguaglianza $(\Delta A)_\psi(\Delta B)_\psi\geq|\langle i[A, B]\rangle_\psi|/2$, dove $[A, B]=AB-BA$. In tal modo, se A e B non commutano sullo stato ψ, cioè $|\langle i[A, B]\rangle_\psi|\neq 0$, allora in tale stato le grandezze fisiche A e B non possono avere simultaneamente un valore esatto: il prodotto delle loro dispersioni è inferiormente limitato. Questa è la relazione d'indeterminazione di Heisenberg. Dato che $[X_j, P_k]=ih\delta_{jk}$ ($\delta_{jk}=1$ per $j=k$ e 0 per $j\neq k$), la j-esima coordinata della particella e la proiezione del suo impulso sul j-esimo asse coordinato non possono avere contemporaneamente valori esatti: $\Delta x_j\Delta p_j\geq h/2$. I motivi intuitivi di ciò (a livello di modelli matematici) sono stati discussi nell'articolo «Continuo/discreto»: una particella con valore definito dell'impulso (cioè ben localizzata nello spazio degli impulsi) ha come funzione ψ un'onda armonica nello spazio delle coordinate, e perciò è assolutamente non-localizzata in esso. Per localizzare una particella nello spazio delle coordinate occorre raccogliere il «pacchetto d'onda», nel quale entrano le componenti armoniche con dispersione delle componenti degl'impulsi (numeri d'onda) e, quanto piú è spazialmente compatto questo pacchetto d'onda, tanto piú forte è questa dispersione.

La circostanza che una funzione e la sua trasformata di Fourier, che si presentano da un certo punto di vista come un unico oggetto, posseggano nello stesso tempo proprietà complementari per quanto riguarda le loro localizzazioni, s'incontra in diversi contesti.

Ad esempio, per una funzione f con supporto compatto in $\mathbf{R}^n$ si definisce la norma $\|f\|_{k,l}^2=\int_{\mathbf{R}^n}|(1+|x|^{2k}+\Delta^l)f|^2\,dx$, dove $\Delta=-\sum_{i=1}^{n}\frac{\partial^2}{\partial x_i^2}$. La chiusura rispetto a questa include le funzioni che hanno $2l$ derivate decrescenti all'infinito come $|x|^{n+2-2k}$. La chiusura, nella topologia corrispondente a tutto il sistema di norme, è lo spazio di Schwartz. La trasformata di Fourier lo trasforma in sé, e

inoltre $\|f\|_{k,l} = \|\hat{f}\|_{l,k}$. Ciò indica che dal punto di vista della trasformata di Fourier risultano complementari le proprietà di regolarità da una parte, e quelle di decrescenza rapida all'infinito (cioè del tipo di localizzazione) dall'altra.

Infine, la trasformata di Fourier si può definire per alcuni spazi di funzioni generalizzate (lentamente crescenti) mediante l'uguaglianza di Parseval. Ponendo $\hat{T}(u) = T(\hat{u})$ (T è una distribuzione, u una funzione fondamentale), si trova, ad esempio: $\hat{\delta}_0 = [1/(2\pi)^{n/2}]T_1$ (T_1 è la distribuzione $u \mapsto \int_{\mathbf{R}^n} u\,dx$, corrispondente alla funzione $f \equiv 1$). Qui il carattere complementare della localizzazione di δ_0 e di $\hat{\delta}_0$ si presenta nella forma estrema: punto/spazio.

Cosí come uno spazio simplettico unifica uno spazio e il suo duale in un unico oggetto, si possono unificare uno spazio di funzioni e quello delle loro trasformate di Fourier in uno spazio di funzioni con un numero di variabili doppio. In quest'ultimo, oltre all'operatore trasformata di Fourier, agiscono operatori piú generali, introdotti da Hörmander. Questo approccio, che include necessariamente anche l'esame della struttura simplettica sullo spazio delle variabili, permette di considerare da un unico punto di vista molte questioni di teoria delle equazioni differenziali alle derivate parziali. [I. M. G. e JU. I. M.].

L'**idea di** dualità ha un carattere estremamente generale, sia nelle **matematiche**, sia in **fisica**, in particolare nella meccanica quantistica (cfr. **quanti**, **particella**), sia in biologia (cfr. **genotipo/fenotipo**). In effetti essa è presente in tutte le **coppie filosofiche**, sia come dualità fenomenologica (ad esempio **puro/impuro**, **dipendenza/indipendenza**, **equilibrio/squilibrio**) sia come presenza/assenza di un contrassegno (ad esempio **uomo/donna**, **salute/malattia**) sia infine come complementarità di carattere metodologico (ad esempio **analisi/sintesi**, **soggetto/oggetto**, **locale/globale**).

In questo articolo sono illustrati gli usi in matematica del termine 'dualità': in **logica** (cfr. anche **insieme**, **assioma/postulato**), in algebra (cfr. **continuo/discreto**, **applicazioni**, **simmetria**, **razionale/algebrico/trascendente**, **divisibilità**), in geometria, in analisi funzionale (cfr. **spettro**, **calcolo**).

Insieme

1. *I problemi fondamentali.*

Se nella matematica contemporanea esiste un concetto «piú importante», senza dubbio si tratta del concetto di insieme. Un accurato esame degli articoli comparsi nelle maggiori riviste matematiche degli ultimi decenni farebbe scoprire che la maggior parte dei concetti matematici piú studiati è costituita da insiemi dotati di una struttura. Le rassegne storiche della matematica del XX secolo sono, d'altro lato, dedicate in misura considerevole al ruolo della teoria degli insiemi nei fondamenti della matematica, ai paradossi della teoria degli insiemi, alla crisi dei fondamenti e alla constatazione dell'assenza di un accordo generale sul modo di uscire da questa crisi.

In effetti la formulazione della teoria degli insiemi nei lavori di Georg Cantor negli anni '70-80 del secolo scorso si è rivelata la causa fondamentale della ristrutturazione di tutta la matematica, oltre che un catalizzatore di profondi e difficili processi della propria autocoscienza. La penetrazione del pensiero insiemistico in tutte le discipline matematiche si è accompagnata con il riconoscimento dell'unità e della forza vitale dell'immenso organismo delle conoscenze matematiche, con la creazione di concetti e di formalismi efficienti non solo in fisica e nella pratica, ma che hanno lasciato una profonda impronta anche sul carattere delle scienze umanistiche. Contemporaneamente queste considerazioni di linguaggio e di semantica della teoria degli insiemi hanno condotto a una interpretazione assolutamente nuova della natura della verità matematica, della dimostrazione, dell'infinito; alla scoperta di problemi non decibibili; a una profonda divisione degli specialisti di logica e di filosofia della matematica in diverse scuole e gruppi.

L'esplicazione del concetto di insieme presenta notevoli difficoltà a causa della sua centralità, della sua irriducibilità a concetti matematici piú elementari. Il metodo per introdurre i concetti fondamentali in matematica non è sostanzialmente cambiato dai tempi di Euclide, costretto a definire «punti» e «rette» prima di formulare gli assiomi della sua geometria. È cambiato soltanto il grado di consapevolezza delle nostre attività, cosa che ha permesso di chiarirne le relazioni dal punto di vista della filosofia della scienza.

Tale metodo consiste nel condurre l'esplicazione parallelamente da due punti di vista fra loro complementari: quello «significativo» o «intuitivo» da una parte, quello «formale» dall'altra.

La definizione intuitiva di insieme come «unione in un tutto di oggetti distinti della nostra intuizione o del nostro pensiero» (Cantor) si appella ad alcuni archetipi fondamentali ed eccessivamente generali del pensiero umano (a cui si riferisce anche la definizione euclidea di punto come «posizione senza lunghezza e larghezza»). Questi archetipi senza dubbio sono suscettibili di ulteriore indagine coi metodi della psicologia e della neurofisiologia (in particolare per ri-

solvere il problema del ruolo relativo delle strutture della conoscenza – innate o acquisite con lo studio –, e della loro formazione). Tuttavia qualsiasi ricerca del genere conduce oltre i confini propri della matematica. Queste definizioni svolgono il ruolo del diapason per accordare le diverse conoscenze individuali sullo stesso tono, ma esse stesse si trovano fuori della matematica, cosa che è inevitabile nell'introduzione dei concetti fondamentali, a meno di assumere in maniera del tutto conseguente un altro punto di vista, quello formale.

Le definizioni formali non sono propriamente considerate *concetti*, ma *termini*. I termini iniziali non sono definiti; solo i contesti nei quali sono usati correttamente sono suscettibili di una descrizione. I termini composti sono abbreviazioni di determinati testi costituiti dai termini iniziali.

Per le applicazioni agli insiemi, ciò significa che occorre fissare il linguaggio nel quale si deve parlare degli insiemi. Questo linguaggio può essere completamente formalizzato, vale a dire costruito per intero in modo artificiale con regole sintattiche rigorose (cfr. § 2), o comprendere frammenti piú o meno grandi di linguaggio naturale (cfr. l'articolo «Assioma/postulato» di questa stessa *Enciclopedia*). In ogni modo devono essere definite in maniera il piú possibile netta, in primo luogo, le regole di formazione delle frasi, includenti i nomi degli insiemi, il predicato «appartenere a» e i termini logici («se ... allora», «e», «o», «non», «per ogni», ecc.), in secondo luogo, le regole di formazione delle successioni di frasi, da considerarsi dimostrazioni matematiche di teoria degli insiemi. La semantica di un tale linguaggio, vale a dire le regole di comprensione del «significato» delle frasi e delle dimostrazioni, in linea di principio, può essere completamente sottaciuta.

Si capisce che le concezioni «intuitiva» e «formale» della teoria degli insiemi (e della matematica in generale) sono in realtà strettamente connesse. I formalismi della teoria degli insiemi sono sorti come risultato della scoperta di contesti apparentemente naturali nei quali l'uso della parola 'insieme', nel senso di Cantor, conduce rapidamente a contraddizioni logiche (si veda la descrizione del paradosso di Russell nel § 5 dell'articolo «Assioma/postulato»). I tentativi di escludere la comparsa di questi contesti per mezzo di divieti puramente linguistici, che conservassero le procedure semiintuitive della versione insiemistica della matematica classica, hanno condotto ad assiomatizzazioni della teoria degli insiemi quali il sistema di Zermelo-Fraenkel, di Neumann, di Gödel-Bernays. Ma la scelta stessa di questi divieti era motivata proprio dalle rappresentazioni intuitive relative alle motivazioni delle difficoltà incontrate (come la descrizione di insiemi definiti «da se stessi» o «troppo grandi», ecc.). Per questo la cristallizzazione dei linguaggi formali si è accompagnata anche a una determinata precisazione della forma intuitiva dell'«insieme», per la quale la definizione di Cantor risultava soltanto il punto storico d'inizio. Purtroppo in questo modo non si è sviluppata un'unica rappresentazione generale per l'insieme matematico e un attento esame del problema fa vedere che è poco probabile che essa possa nascere.

La consapevolezza della suddivisione della matematica in «testo» e «significato» è il frutto dell'introduzione del programma di Hilbert come base per

la matematica. Questo programma non serví per la soluzione dei problemi posti da Hilbert, ma presto permise di scoprire che essi erano mal posti.

Secondo Hilbert il ruolo del formalismo doveva consistere nel garantire l'attendibilità delle considerazioni matematiche, per mezzo della verifica della loro non-contraddittorietà formale, indipendentemente dal contenuto. Questa verifica, per ipotesi, doveva svolgersi con metodi piú attendibili (finitistici) di quelli usati dalla teoria intuitiva degli insiemi. Tuttavia Gödel dimostrò che lo studio del formalismo intuitivo è un problema matematico di per sé tanto difficile che nella dimostrazione della sua non-contraddittorietà non è sufficiente usare neppure i metodi descrivibili con questo formalismo; e in ogni caso i metodi finitistici non possono bastare.

In un ambito piú ampio il ruolo del formalismo potrebbe almeno consistere nel contribuire al chiarimento dell'intuizione necessaria a garantire una comprensione unitaria dell'«insieme» e nel fondare la verità sulla non-contraddittorietà della teoria degli insiemi, almeno nella stessa misura in cui ciò è corretto per il concetto di numero e di aritmetica elementare. In qualche misura questo scopo è stato raggiunto. Nicolas Bourbaki e le sue idee, accettate dalla grande comunità matematica internazionale, dimostrano sia l'estesa diffusione di una unità di punti di vista sugli insiemi, sia una grande produttività dei principî insiemistici. In maniera grossolana si può dire che nella pratica è stato dimostrato che, entro limiti noti e abbastanza ampi, la non-determinatezza del concetto stesso di insieme non altera la comprensione unitaria, la produzione, lo scambio e la valutazione dei testi matematici, nei quali svolge un ruolo fondamentale. Tuttavia non si possono chiudere gli occhi di fronte alla reazione di gruppi di matematici (intuizionisti, costruttivisti) poco numerosi ma con seri argomenti, i quali mettono in guardia sui pericoli possibili lungo questa strada. Queste argomentazioni verranno riprese nell'ultimo paragrafo di questo articolo.

Sotto queste condizioni l'assunzione di un punto di vista unitario sugli insiemi conduce a una netta unificazione dell'intuizione, della tecnica e delle possibilità potenziali della matematica in quanto scienza. La stessa cosa è senz'altro valida per l'esplicazione di qualsivoglia sistema alternativo di concetti, che possa essere posto a fondamento della matematica in futuro. Si ritiene che il carattere di complementarità (nel senso di Bohr) degli aspetti intuitivi e formali delle teorie matematiche sia una delle principali lezioni della metamatematica del XX secolo. È parallelo alle analoghe lezioni gnoseologiche della microfisica e della biologia. La profonda dualità dei concetti di testo e significato manifestatasi inizialmente nei modelli matematici sia di testi (linguaggi formali) sia di significati (le loro interpretazioni), ha senza dubbio un valore generale. Il significato non può essere esplicato e oggettivato al di fuori del testo; ma il testo non è in grado di trasmettere né «tutto il significato», né «solo questo significato». In matematica ciò è vero piú che in altre scienze, i cui testi hanno spesso un significato pragmatico esterno a loro. Una comprensione piú profonda sia della semantica sia dei problemi dei fondamenti della matematica può essere raggiunta ampliando, e non restringendo, i fatti e i metodi in considerazione. La matematica non è un gioco puramente formale, né un prodotto della pura intuizione individuale; è un aspetto dell'attività della comunità delle coscienze individuali nel mondo esterno, e solo tenendo conto di una quantità di fattori maggiore di quella considerata finora si può giungere alla costruzione di un modello verosimile della matematica, nel quale gli aspetti intuitivo e formale risultino elementi di una stessa unità.

Tornando al linguaggio della teoria degli insiemi (considerato in dettaglio piú oltre, come pure negli articoli «Assioma/postulato», «Applicazioni», «Strutture matematiche», che con questo articolo formano un complesso unitario) è opportuno indicare subito i tratti caratteristici di questo linguaggio, i quali l'hanno proprio posto al centro delle considerazioni precedenti. Dal nostro punto di vista sono la sua economicità, l'universalità e «l'accesso diretto» ai problemi dell'infinito.

L'economicità dei metodi espressivi del linguaggio della teoria degli insiemi si manifesta nell'uso, oltre che dei metodi logici, di due soli predicati non definiti, «essere un insieme» e «appartenere a». All'inizio (in particolare con Cantor) a questi si aggiungeva un numero non grande di predicati del tipo «essere una funzione» oppure «essere una relazione», ma in seguito si dimostrò come anche questi si riducano ai due originali (cfr. § 3). Grazie alla economicità del linguaggio degli insiemi, la sua formalizzazione graduale e la sua analisi logica furono comode pietre di paragone per la formazione di metodi di ricerca metamatematica (fra l'altro questa stessa funzione è condivisa col linguaggio dell'aritmetica elementare, la cui superiorità consiste in una trasparenza della semantica notevolmente maggiore).

L'universalità del linguaggio della teoria degli insiemi si è manifestata prima di tutto nella rapida formulazione di tecniche di traduzione di tutti i concetti della matematica classica in questo linguaggio, a seguito della quale si scoprí la loro unità interna, mascherata da una varietà di termini e d'intuizioni delle epoche precedenti. Il successivo sviluppo della matematica ha dimostrato l'eccezionale comodità del linguaggio degli insiemi, congiunto a variabili metalinguistiche del tipo «strutture» di Bourbaki, per collegare, scoprire e precisare nuovi fatti matematici. Oltre a ciò il linguaggio degli insiemi si rivelò tanto adattabile da accogliere in sé un intero spettro d'interpretazioni di approcci alternativi ai concetti principali della matematica, spesso contrari, nello spirito, alla teoria degli insiemi. Ad esempio tutta la matematica costruttiva, eretta intorno al concetto di algoritmo, senza alcuna difficoltà «si immerge in maniera isomorfa» nella matematica insiemistica, sebbene i costruttivisti contestino la sensatezza di questa immersione. Lo stesso vale per la logica intuizionista.

Infine i problemi di Cantor sull'infinito attuale hanno svolto in matematica un ruolo veramente particolare. «L'infinità nuda» di elementi distinguibili ma non aventi alcuna altra individualità, secondo Cantor, ha aperto alla coscienza matematica una serie di procedimenti concettuali di valore unicò, quali il procedimento di diagonalizzazione, l'induzione transfinita, l'assioma della scelta. Questi procedimenti conservano il loro significato anche sotto interpretazioni nettamente diverse. La possibilità stessa di considerare matematicamente significativa l'infinità attuale, liberandola sia da strutture addizionali non

essenziali, sia dalla barriera psicologica di due millenni, è stata una scoperta di grandissima importanza che ha valore per tutte le scienze. In qualunque modo si valuti il risultato del successivo periodo di dura critica alle «teorie dell'infinito», la dimostrazione della loro stessa esistenza è un momento cruciale nella storia della scienza, le cui lezioni sono imperiture.

2. *Teorie formali degli insiemi.*

Nell'articolo «Assioma/postulato» sono stati descritti a livello qualitativo il linguaggio della teoria degli insiemi, i suoi assiomi e la semantica; nell'articolo «Applicazioni» sono state riportate alcune costruzioni fondamentali della teoria degli insiemi. Per questo si comincia qui con l'introduzione del linguaggio formale della teoria degli insiemi.

La descrizione di un linguaggio formale L si riduce in generale alle seguenti componenti: *a*) la descrizione dell'alfabeto A del linguaggio: l'insieme dei suoi simboli; *b*) la descrizione delle regole di formazione di particolari successioni finite di simboli: le espressioni «sensate». Queste descrizioni solitamente si accompagnano con convenzioni relative ad abbreviazioni di scrittura (senza le quali i testi nel linguaggio formale diverrebbero rapidamente lunghi in maniera eccessiva), con commenti relativi a equivalenze verbali di enunciati formali e con la spiegazione del loro significato intuitivo; *c*) la descrizione delle regole di formazione di testi particolari – successioni finite di espressioni sensate, le quali costituiscono le deduzioni formali, vale a dire i modelli formali delle dimostrazioni matematiche. Queste regole di formazione possono essere scelte in diversi modi; uno dei metodi essenziali per variarle è assumere questi o quegli assiomi formali, queste o quelle regole di deduzione. Ogni scelta delle regole di deduzione formale definisce una teoria formale sul linguaggio L.

Si descriverà ora il particolare linguaggio del primo ordine della teoria degli insiemi L_1 Set, secondo l'assiomatica di Zermelo-Fraenkel.

2.1. Alfabeto del linguaggio della teoria degli insiemi L_1 Set.

Si suddivide nei seguenti gruppi:

a) un insieme numerabile di simboli per le variabili $x, y, z, \ldots$ dotati di indici;
b) il simbolo dell'insieme vuoto $\varnothing$;
c) il simbolo della relazione «appartenere a» $\in$;
d) il simbolo della relazione di uguaglianza $=$;
e) i simboli dei connettivi logici e dei quantificatori, $\leftrightarrow$ (equivalenza), $\rightarrow$ (implicazione), $\vee$ (o non esclusivo), $\wedge$ (e), $\neg$ (non), $\forall$ (quantificatore universale), $\exists$ (quantificatore esistenziale);
f) le parentesi «(» e «)».

2.2. Formule del linguaggio L_1 Set.

Le espressioni sensate del linguaggio L_1 Set si chiamano formule. L'insieme delle formule è definito mediante le seguenti regole di formazione:

a) se X, Y sono due variabili oppure $\varnothing$, e se $*$ è uno qualunque dei simboli $=$, $\in$, allora $X*Y$ è una formula;
b) se P e Q sono formule e se x è una variabile, allora $(P)\leftrightarrow(Q)$, $(P)\rightarrow(Q)$, $(P)\vee(Q)$, $(P)\wedge(Q)$, $\neg(P)$, $\forall x(P)$, $\exists x(P)$ sono formule;
c) l'insieme delle formule è esaurito dalle espressioni che si possono costruire applicando le regole *a*) e *b*).

Le formule descritte dalle regole *a*) si dicono *atomiche*. Ogni formula che non sia atomica può essere univocamente rappresentata esattamente in una delle sette forme descritte dalle regole *b*). Le componenti P, Q che entrano nella corrispondente rappresentazione di una formula originaria possono essere a loro volta sia atomiche sia rappresentabili in una di queste forme. Questo procedimento di analisi sintattica si svolge meccanicamente e permette di costruire ogni formula a partire da quelle atomiche. Le definizioni fondamentali e i teoremi metalinguistici relativi alle formule di L_1 Set si stabiliscono per induzione sulla lunghezza delle formule, a partire da quelle atomiche.

Un esempio importante di questo procedimento è la definizione di occorrenze libere e vincolate delle variabili in una formula. Tutte le occorrenze delle variabili in una formula atomica sono considerate libere. Nelle formule $\forall x(P)$, $\exists x(P)$ tutte le occorrenze della variabile x sono considerate vincolate, mentre le occorrenze delle altre variabili dipendono da P. Nelle rimanenti cinque formule della regola *b*) le occorrenze delle variabili sono considerate le stesse delle occorrenze a loro corrispondenti nelle sottoformule P, Q. La variabile x è libera in P se almeno una delle sue occorrenze è libera in P.

La traduzione verbale delle espressioni del linguaggio L_1 Set si realizza per mezzo delle seguenti regole non formali:

1) i simboli di variabile $x, y, z, \ldots$ sono nomi di insiemi indeterminati; il simbolo $\varnothing$ è il nome dell'insieme vuoto;
2) la formula atomica $x\in y$ si legge «L'insieme x appartiene all'insieme y» o in qualche maniera sinonima; la formula atomica $x=y$ si legge «L'insieme x è uguale all'insieme y»;
3) la lettura delle formule $(P)\leftrightarrow(Q), \ldots, \exists x(P)$ deve risultare chiara dalle denominazioni dei simboli dei connettivi e dei quantificatori. In tal modo le formule sono scritture di enunciati relativi a insiemi indeterminati.

Poiché la sintassi del linguaggio naturale non è normalizzabile tanto nettamente quanto la sintassi di L_1 Set e non ammette catene troppo lunghe di frasi connesse l'una all'altra, la traduzione richiede delle abbreviazioni, delle approssimazioni espressive, ecc.

Si dànno qui alcuni esempi di formule e di loro traduzioni:

$(\forall z((z \in x) \leftrightarrow (z \in y))) \leftrightarrow (x = y)$: «Gli insiemi x e y sono uguali se e solo se ogni insieme z che appartiene a x appartiene anche a y e viceversa». Si tratta della scrittura in linguaggio formale dell'assioma di estensionalità (cfr. l'articolo «Assioma/postulato», § 4, in questa stessa *Enciclopedia*).

$\neg(\exists x(x \in \emptyset))$: «Non è vero che esiste x appartenente all'insieme vuoto».

È naturale considerare la formula P come un enunciato precisamente relativo a quegli insiemi i cui simboli sono liberi in P. I simboli vincolati dai quantificatori $\forall$ ed $\exists$ svolgono il ruolo di variabili ausiliarie nella formula considerata. Con l'introduzione del concetto di modello, o interpretazione di un linguaggio formale, alcune formule, fra cui tutte quelle chiuse, vale a dire non contenenti variabili libere, risultano univocamente determinate per quanto riguarda la loro verità: esse sono vere o false. In generale, per le formule contenenti variabili libere risulta determinata la verità solo dopo che a queste variabili è assegnato un determinato valore (per maggiori particolari cfr. § 4).

2.3. Assiomi particolari formali del linguaggio L_1 Set.

Tutti gli assiomi di Zermelo-Fraenkel (ZF1-ZF10) riportati nel § 4 dell'articolo «Assioma/postulato» possono essere tradotti nel linguaggio formale, analogamente a quanto è stato fatto per l'assioma di estensionalità e per l'assioma dell'insieme vuoto. Le formulazioni verbali degli assiomi ZF1-ZF10 sono state scelte in modo da facilitare al massimo la possibilità di questa traduzione al lettore interessato.

Due casi richiedono commenti a parte.

In primo luogo negli assiomi ZF7-ZF9 figurava il concetto di «proprietà dell'insieme degli z». Occorre sostituirlo con il concetto di «formula del linguaggio L_1 Set nella quale solo la variabile z compare libera». In altri termini si riduce il concetto di «proprietà di un insieme» al concetto di «proprietà esprimibile nel linguaggio L_1 Set». Poiché nel linguaggio L_1 Set non ci sono metodi per tradurre l'espressione «insieme qualsiasi» (i quantificatori $\forall P$ sulle formule variabili P non sono ammessi), l'assioma ZF7 si trasforma in uno «schema di assiomi formali», un assioma formale per ogni formula $P(z)$ con la variabile libera z, della forma

$$\forall x \exists y \forall z(z \in y \rightarrow (z \in x \wedge P(x))).$$

In questa scrittura, come è accettato, sono state omesse tante parentesi quante può ricostruire senza troppa difficoltà il lettore attento. In maniera analoga, solo leggermente piú complicata, si tratta l'assioma ZF8, nel quale compaiono le proprietà delle coppie ordinate di insiemi. Questa precisazione del significato da attribuire al concetto di «proprietà» è il primo apporto essenziale della formalizzazione del linguaggio nella teoria intuitiva degli insiemi. Raggiungere questo scopo con metodi essenzialmente diversi sarebbe stato alquanto difficoltoso. D'altro lato questo scopo è raggiunto a caro prezzo: la limitazione

del numero di proprietà che si possono considerare. Soltanto un continuo processo di traduzione della matematica nel linguaggio L_1 Set può mostrare che le proprietà esprimibili in L_1 Set sono sufficienti.

In secondo luogo nella formulazione degli assiomi ZF1-ZF10 si sono spesso usate abbreviazioni per designare insiemi la cui esistenza era assicurata da assiomi: $\{x, y\}$, $\mathcal{P}(x)$, ecc. È comodo ottenere in due passi la traduzione delle espressioni verbali in cui compaiono queste denominazioni in linguaggio formale. Al primo passo si costruisce una «quasiformula» per mezzo di $\in$, $=$, i connettivi e i quantificatori, nella quale possono rientrare sia le variabili sia i «termini» del tipo $\{x, y\}$ o $\mathcal{P}(x)$. Al secondo passo ogni termine di questo tipo viene sostituito da una nuova variabile, ad esempio z, e per z si dà una corrispondente «definizione» con variabili ausiliarie. Ad esempio la «quasiformula» $\mathcal{P}(x) \in y$ si trasforma nella formula

$$(z \in y) \wedge \forall v(v \in z \leftrightarrow \forall w(w \in v \rightarrow w \in y)).$$

Negli studi sulla teoria formale degli insiemi (e in generale sui linguaggi formali) abbreviazioni di scrittura come l'uso di quasiformule si producono in continuità.

2.4. Assiomi logici del linguaggio L_1 Set.

Fanno parte di questi alcune serie di formule chiaramente descrivibili, ad esempio le tautologie, come $P \rightarrow (Q \rightarrow P)$ oppure $\neg\neg P \leftrightarrow P$, assiomi con quantificatori come $\forall x \neg P \leftrightarrow \exists x P$ e $\forall x P \rightarrow P(y)$ (dove $P(y)$ è la scrittura risultante dalla sostituzione del simbolo y al posto di tutte le occorrenze libere del simbolo x nella formula P), e infine assiomi di uguaglianza, come $x = y \rightarrow (P(x) \rightarrow P(y))$.

Gli assiomi logici sono equivalenti formali di espressioni di significato generale, la cui verità, di fatto, non dipende dall'interpretazione delle loro componenti.

2.5. Deduzioni formali nel linguaggio L_1 Set.

Sia $\mathfrak{E}$ un insieme di formule del linguaggio L_1 Set. Si dice deduzione formale da $\mathfrak{E}$ una successione di formule $P_1, \ldots, P_n$ del linguaggio che abbia le seguenti proprietà: per ogni $i \leq n$, $P_i \in \mathfrak{E}$; oppure P_i è della forma $\forall x P_j$ con $j < i$; oppure fra le formule $P_1, \ldots, P_{i-1}$ c'è una formula del tipo $P_k \rightarrow P_i$ con $k \leq i-1$.

In altri termini, ogni formula P_i della deduzione è una formula di $\mathfrak{E}$ oppure è una conseguenza immediata della formula precedente P_j mediante la regola GEN (generalizzazione), oppure è una conseguenza immediata di formule precedenti P_k, $P_k \rightarrow P_i$ mediante la regola MP (*modus ponens*).

La deduzione formale $P_1, \ldots, P_n$ «deduce» da $\mathfrak{E}$ l'ultima formula P_n.

Se $\mathfrak{E}$ è la totalità degli assiomi particolari e degli assiomi logici del linguaggio L_1 Set, allora le sue deduzioni formali sono corrispondenti alle dimostra-

zioni di teoria degli insiemi: ogni deduzione formale «dimostra» la sua ultima formula. La totalità delle formule deducibili da un dato sistema di assiomi costituisce (una delle) teorie formali degli insiemi.

2.6. Proprietà collettive ed estensioni del linguaggio L_1 Set.

Come è stato osservato in «Assioma/postulato» quasi tutti gli assiomi di Zermelo-Fraenkel postulano l'esistenza di un insieme x costruito a partire da alcuni insiemi dati $y_1, ..., y_r$ per mezzo di una qualche costruzione. Oltre a ciò, i piú importanti sono i casi in cui x è definito univocamente a partire da $y_1, ..., y_r$ indicando che tutti gli elementi $y \in x$ devono possedere insieme a $y_1, ..., y_r$ una certa proprietà P. Di questo tipo sono l'assioma della coppia ZF3 (ove P è $y = y_1 \vee y = y_2$) l'assioma dell'unione ZF4 (ove P è $\exists z(y \in z \wedge \wedge z \in y_1)$) e altri.

Nella teoria formale la formula $P(y; y_1, ..., y_r)$ si dice *collettiva* per y se in questa teoria è deducibile la formula $\exists x(y \in x \leftrightarrow P(y; y_1, ..., y_r))$. Nella teoria intuitiva degli insiemi a ogni proprietà collettiva P si può far corrispondere l'operazione di riunione degli insiemi y con la proprietà P; il risultato di questa riunione è l'insieme usualmente denotato $\{y \mid P(y; y_1, ..., y_r)\}$. Per proprietà P particolarmente importanti possono essere scelte speciali denominazioni, come $\{y_1, y_2\}$ (la coppia) oppure $\mathcal{P}(x)$ (le parti di x). È stato spiegato sopra come sia possibile fare a meno di questi insiemi «composti» o «termini», nella traduzione, nel linguaggio L_1 Set, di espressioni semiverbali contenenti questi termini. È tuttavia possibile anche un'altra strada: quella di estendere il linguaggio formale aggiungendo alle espressioni ammissibili: *a*) termini per le operazioni, definiti induttivamente, rispondenti a qualche proprietà collettiva della data teoria; *b*) tutte le formule atomiche del tipo $t_1 \in t_2$, $t_1 = t_2$, in cui si trovino termini, vale a dire tutti i nomi degli insiemi; *c*) tutte le formule ottenibili da quelle atomiche per mezzo dei connettivi logici e dei quantificatori. Aggiungendo all'elenco degli assiomi particolari della teoria le formule che definiscono i termini corrispondenti ed estendendo alquanto l'elenco degli assiomi logici per includervi gli assiomi con termini, si ottiene un'estensione di questa teoria in un linguaggio ampliato. Si può dimostrare un teorema del tutto naturale sul fatto che, nella teoria estesa, sono dimostrabili solo le formule equivalenti a formule dimostrabili nella vecchia teoria, di modo che queste estensioni di linguaggio sono inessenziali da un punto di vista logico e vi si può ricorrere se ciò risulta comodo da un punto di vista tecnico.

2.7. Problemi di metateoria delle teorie formali degli insiemi.

Nel § 3 di «Assioma/postulato» è stato dato un elenco dei concetti fondamentali della metateoria delle teorie assiomatiche, con brevi commenti. Nell'applicazione alle teorie formali questi concetti sono sottoposti a una notevole precisazione e i problemi ad essi collegati diventano questioni formulabili in maniera precisa dal punto di vista matematico.

Cosí la *non-contraddittorietà* formale degli assiomi di Zermelo-Fraenkel significa che non esiste una formula P nel linguaggio L_1 Set tale che P e $\neg P$ siano contemporaneamente deducibili dagli assiomi formali (affermazione equivalente: non tutte le formule di L_1 Set sono deducibili dagli assiomi formali). Sebbene il problema di determinare la non-contraddittorietà, in questa forma, sia un problema matematico relativo a successioni di oggetti combinatori costruibili, esso risulta molto difficile. Una qualche argomentazione specifica a sostegno di questa non-contraddittorietà, tranne la fiducia nella validità dell'intuizione per la scelta del sistema di assiomi (confermata dallo sviluppo della matematica nel corso di alcuni decenni), non esiste ancora. D'altro lato, dal tempo dei primi lavori di Hilbert, è mutata notevolmente anche la valutazione dell'importanza di questi problemi. La non-contraddittorietà formale è solo un rozzo modello di alcune caratteristiche abbastanza sottili delle teorie matematiche significative, e si può realmente dubitare del valore gnoseologico delle deduzioni ottenute per simili modelli. Si tornerà su questo problema nell'ultimo paragrafo di questo articolo.

Un serio progresso si è raggiunto nel problema della *deducibilità* di varie classi di formule dagli assiomi di Zermelo-Fraenkel. Una pietra di paragone decennale per i metodi metateorici, in questo campo, è stata costituita dal problema del continuo di Cantor (cfr. l'articolo «Applicazioni», § 2, in questa stessa *Enciclopedia*). Recentemente è stato dimostrato che la formula IC nel linguaggio L_1 Set, la quale è una traduzione dell'espressione «Esiste un insieme di potenza propriamente intermedia fra il numerabile e il continuo», non è deducibile dagli assiomi di Zermelo-Fraenkel, come pure la negazione di questa formula (sotto la condizione che gli assiomi stessi di Zermelo-Fraenkel siano formalmente non-contraddittori).

I metodi di dimostrazione dell'indipendenza di formule come IC dagli assiomi comprendono uno studio minuzioso delle diverse interpretazioni del linguaggio L_1 Set; qualcosa su ciò sarà detto nel § 4. Soltanto in termini di queste interpretazioni si è cominciato a parlare di verità (piuttosto che deducibilità) nelle teorie formali degli insiemi, la quale, in tal modo, ha un carattere incomparabilmente piú indeterminato e relativo che, ad esempio, la verità in aritmetica. Si può dire che le ricerche sull'ipotesi di Cantor abbiano esaurito il dubbio relativo al significato dell'impostazione stessa del problema della sua verità o falsità. Sebbene simili dubbi siano stati espressi nel corso di tutta la sua storia, il progresso consiste nel fatto che ora essi sono argomentati molto meglio; per questo motivo le obiezioni contrarie (ad esempio le proposte di nuovi assiomi della teoria degli insiemi) possono essere formulate in forma piú concreta. La formalizzazione della vecchia intuizione ha detto, su questo problema, la sua ultima parola. Ai sostenitori della teoria intuitiva degli insiemi occorre scegliere fra la necessità di sviluppare l'intuizione fino al punto in cui i problemi relativi all'assunzione di nuovi assiomi ricevano una soluzione unanime (cosa poco probabile) e la possibilità che esistano sistemi alternativi per la teoria degli insiemi, che ammettano, in particolare, soluzioni diverse del problema dell'infinito.

3. *Matematica insiemistica.*

Quando si afferma che la matematica ha oggi, in larga parte, un carattere insiemistico, vengono tenute presenti alcune circostanze, connesse ma eterogenee.

Prima di tutto la possibilità di tradurre nel linguaggio dei concetti della teoria degli insiemi e quindi nel linguaggio formale L_1 Set qualunque concetto, costruzione e affermazione relativa a sottoteorie matematiche diverse. Quindi la possibilità di ricondurre le dimostrazioni con metodi formali agli assiomi di Zermelo-Fraenkel (a volte completati da assiomi dell'infinito di un tipo piú forte, ad esempio l'esistenza di ordinali inaccessibili; fra l'altro questi teoremi solitamente sono formulati come implicazioni). La dimostrazione di questa possibilità richiede un grosso lavoro nelle applicazioni all'analisi classica, la «nuova matematica» stessa ormai si formula in termini insiemistici, senza però l'esplicazione costante del concetto di numero reale. Si vedano a questo riguardo gli *Eléments de mathématique* di Bourbaki e alcune osservazioni piú avanti.

Fra le conseguenze a cui questa possibilità ha dato vita, la matematica ha individuato un metodo – iniziato con l'aritmetizzazione dell'analisi – per superare la separazione fra continuo e discreto. Rimuovendo l'originale contrapposizione intuitiva continuo/discreto per mezzo della riduzione di entrambi all'intuizione degli insiemi, la teoria degli insiemi ha poi permesso di costruire un immenso mondo gerarchico di strutture, manifestanti «gradi diversi» di continuità, una struttura continua del discreto (le topologie m-adiche dei numeri interi), invarianti discreti del continuo (topologia algebrica e omotopia), ecc. (cfr. l'articolo «Continuo/discreto» in questa stessa *Enciclopedia*).

In seguito l'ideologia insiemistica ha condotto a una diffusione di una speciale forma *estensionale* di pensiero. Invece di descrizioni in termini di proprietà si assumono descrizioni in termini di insiemi di oggetti con date proprietà. I formalismi concreti sono scelti secondo la possibilità di eliminare la dipendenza dei nomi degli oggetti dalle scelte derivate (sistemi di coordinate, sistemi di numerazione, rappresentanti di classi di equivalenza, ecc.), è il principio di «scrittura invariante» dei matematici o «scrittura covariante» dei fisici. Tutte le possibilità sono da ritenersi attualizzabili (ad esempio in alcuni calcoli di fisica teorica la dimensione dello spazio-tempo è considerata un qualunque numero complesso (!)).

L'esistenza va intesa nel senso piú ampio della parola: non si presuppone né una dimostrazione di esistenza di un oggetto né la sua costruzione effettiva e neppure la sua esprimibilità in un linguaggio abbastanza povero.

Esaminando piú dettagliatamente il problema della traduzione delle diverse «matematiche» nel linguaggio della teoria degli insiemi, si constata che praticamente esso non si pone per quelle nuove discipline come la topologia o l'algebra astratta, che fin dall'inizio sono state formulate entro l'ambito insiemistico. Per questo ci si può limitare a una breve descrizione insiemistica «dei numeri e delle figure», che costituiscono degli argomenti fondamentali della matematica classica, e alla considerazione dei problemi posti dall'uso del linguaggio delle categorie.

3.1. Numeri naturali.

Il modello standard dello zero nella teoria degli insiemi è l'insieme vuoto $\varnothing$; dell'unità $\{\varnothing\}$, del due $\{\varnothing, \{\varnothing\}\}$; in generale se l'insieme $\tilde{n}$, corrispondente al numero n, è già stato definito, allora $(n+1)^{\sim} = \tilde{n} \cup \{\tilde{n}\}$. L'insieme $\tilde{n}$ è un ordinale finito con n elementi: esso è totalmente ordinato dalla relazione $\in$ fra i suoi elementi (cfr. § 4). L'esistenza di ciascun numero naturale è garantita dagli assiomi di Zermelo-Fraenkel, in particolare ZF_3 e ZF6. In L_1 Set non è possibile scrivere i termini che li designano, ma si possono utilizzare espressioni approssimative, oppure estendere L_1 Set, come è stato spiegato nel § 2. Poiché l'operazione $n \to n+1$ si esprime nel linguaggio della teoria degli insiemi, in questo stesso linguaggio si esprimono tutte le operazioni aritmetiche sui numeri naturali, grazie alla definizione induttiva dell'aritmetica di Peano. L'assioma di induzione è dimostrabile a partire dagli assiomi di Zermelo-Fraenkel.

3.2. Numeri relativi e razionali.

La riduzione classica dei numeri relativi ai naturali consiste nel considerare classi di equivalenza di coppie ordinate di numeri naturali in base alla relazione di equivalenza $\langle m, n\rangle \sim \langle m', n'\rangle$ se $m+n' = m'+n$; la classe della coppia $\langle m, n\rangle$ corrisponde al numero $m-n$. Questa classe, ovviamente, è un insieme descrivibile con una formula di L_1 Set; essa può anche essere assunta come definizione del corrispondente numero relativo. Un lavoro abbastanza meticoloso è richiesto per definire le operazioni di somma e di moltiplicazione e per dimostrare le loro proprietà, ma in linea di principio è chiaro come ciò vada fatto (cfr. anche l'articolo «Calcolo», § 3.1, in questa stessa *Enciclopedia*).

Analogamente un numero razionale è costruito come una classe di coppie ordinate $\langle a, b\rangle$ di numeri interi, con $b \neq 0$, secondo la relazione di equivalenza $\langle a, b\rangle \sim \langle a', b'\rangle$ se $ab' = a'b$, dopo di che si definiscono le leggi di composizione e la relazione d'ordine.

3.3. Numeri reali **R**.

Si possono costruire col procedimento di Dedekind (sezioni) o col metodo di Cantor (classi di equivalenza di successioni fondamentali di numeri razionali), i quali hanno un evidente carattere insiemistico e in tal modo sono anche presentati nei corsi istituzionali di analisi (una successione è una funzione dei numeri naturali nei razionali, oppure il suo grafico, vale a dire un insieme).

Vale la pena di osservare che ad ogni gradino di questa scala di costruzioni gli oggetti costruiti al gradino precedente non s'identificano con gli oggetti

costruiti in seguito, ma si applicano soltanto canonicamente in essi. Se si desidera considerare che ogni numero naturale «sia» allo stesso tempo reale, cioè che gli corrisponda uno ed un solo insieme, alla fine della costruzione occorre cambiare la definizione di numero naturale. Allora, ad esempio, all'unità non corrisponderà l'ordinale $\{\varnothing\}$, ma un insieme infinito «di successioni fondamentali di coppie ordinate di ordinali finiti» descrivibile con una formula eccezionalmente lunga di L_1 Set.

In altri termini l'aritmetica si può immergere immediatamente nella teoria degli insiemi, oppure mediante l'analisi; i risultati saranno isomorfi, ma diversi. Con questo esempio sono chiari i motivi per cui le versioni insiemistiche di tutte le discipline matematiche non sono univoche.

3.4. «Figure».

È piú facile immergere la classica geometria euclidea nella teoria degli insiemi, identificando il piano euclideo, munito di un sistema di coordinate cartesiane, con $\mathbf{R}^2 = \mathbf{R} \times \mathbf{R}$. Dopo aver stabilito la possibilità di esprimere la relazione di ordine e le operazioni standard di $\mathbf{R}$ in L_1 Set, è possibile sostituire la geometria euclidea con quella analitica, immersa nella teoria degli insiemi.

L'assiomatica della geometria secondo Hilbert è ovviamente coerente nell'ottica della teoria degli insiemi.

Anche le teorie geometriche contemporanee, che lavorano con spazi topologici, metrici, anellati, ecc., li immergono subito nel linguaggio insiemistico. Se inoltre si usano metriche riemanniane o modelli locali del tipo di spazi euclidei, allora la loro interpretazione insiemistica si riconduce direttamente all'interpretazione del continuo reale.

Una posizione a parte occupa il linguaggio delle categorie, che diventa sempre piú intuitivo e popolare (cfr. l'articolo «Applicazioni», § 6, in questa stessa *Enciclopedia*). Come si è osservato in tale articolo, non esiste una formalizzazione del linguaggio delle categorie che sia completamente indipendente dal linguaggio della teoria degli insiemi. D'altro lato l'interpretazione insiemistica di costruzioni categoriali realmente utilizzate s'imbatte in due difficoltà caratteristiche.

La prima è che non c'è ragione per considerare subito come insiemi gli oggetti e i morfismi di una categoria generale; essi hanno una natura indeterminata. In una intuizione insiemistica e in una formalizzazione adeguata occorre per questo prevedere degli «atomi» – gli oggetti, che non sono insiemi, ma la cui «collezione» costituisce un insieme. L'assiomatica originaria di Zermelo si riferiva esattamente a questo caso, ma sembra che non sia stata perseguita a sufficienza per ottenere tutte le deduzioni categoriali «intuitive».

La seconda difficoltà, probabilmente legata alla prima, consiste nel deciso aumento del ruolo delle «classi» di Neumann-Gödel-Bernays rispetto agli insiemi abituali.

In senso intuitivo una classe è una collezione arbitraria di insiemi, ad esempio «tutti» gli insiemi oppure «tutti» gli ordinali (per chiarimenti piú precisi

sugli universi si veda il paragrafo successivo). Una classe, nel linguaggio formale, si descrive con una «proprietà», o una formula avente una variabile libera. Intuitivamente le classi si distinguono dagli insiemi per il fatto che su esse non si possono eseguire le operazioni possibili sugli insiemi; in primo luogo una classe non può essere un elemento di un'altra classe o di un insieme (sono formalmente vietate le espressioni linguistiche di questo tipo). L'inosservanza di tale divieto conduce immediatamente in prossimità del paradosso di Russell.

Fra l'altro le costruzioni categoriali piú intuitive si riferiscono spesso e volentieri alle classi. Ad esempio la tipica costruzione del «gruppo K_0» di Grothendieck consiste in quanto segue. Si consideri il gruppo libero generato da classi $[A]$ di gruppi abeliani finiti A, a meno di isomorfismi, e lo si quozienti rispetto al sottogruppo generato dalle relazioni $[B]-[A]-[C]$ per tutte le sequenze esatte $0 \to A \to B \to C \to 0$. Questo è il gruppo $K_0(\mathcal{F}\mathcal{A}b)$ per la categoria $\mathcal{F}\mathcal{A}b$ dei gruppi abeliani finiti. I suoi elementi sono classi di classi; nelle assiomatizzazioni standard della teoria degl'insiemi, simili costruzioni sono assolutamente proibite.

Analoghi problemi fondazionali sorgono, ad esempio, nella realizzazione di costruzioni come lo «spazio classificante» di una categoria. In questa costruzione, in maniera semplice, la categoria è paragonata a un grafo infinito i cui vertici sono oggetti e i cui lati sono i morfismi. Dopo ciò, alcuni cammini lungo le frecce di questo grafo, secondo regole determinate, s'incollano in simplessi a piú dimensioni. I gruppi di omologia e di omotopia di questi spazi sono oggetti importanti, spesso già di tipo finito, ma i loro elementi sono ancora classi e classi di classi («gruppi d'ordine superiore K_i»).

La risposta ufficiale dei testi contemporanei, relativa alle difficoltà qui sollevate, consiste nel presupporre che la categoria originaria sia «piccola», vale a dire che i suoi oggetti e i morfismi costituiscano un insieme, e non una classe, all'interno di un universo piú ampio. Le difficoltà allora si riconducono al problema dell'esistenza di una gerarchia di universi insiemistici con le proprietà necessarie (la gerarchia è necessaria per utilizzare liberamente in seguito gli invarianti calcolati, in nuove costruzioni) e al problema dell'indipendenza dei risultati delle costruzioni intuitive dalla scelta di questa gerarchia. Poiché molti di questi problemi sono probabilmente indecidibili, nello stesso senso, ad esempio, dell'ipotesi del continuo, essi sono semplicemente circondati dal silenzio [cfr. Artin, Grothendieck e Verdier 1963-64].

4. *Problemi dell'infinito e modelli della teoria degli insiemi.*

In questo paragrafo sono descritti il concetto di universo e alcuni universi, utilizzati per dimostrare l'indecidibilità di problemi quali l'ipotesi del continuo. In termini semplici, un universo è una classe di insiemi nella quale sono verificati tutti gli assiomi di una teoria assiomatica degli insiemi (nel nostro caso la teoria di Zermelo-Fraenkel); si chiama anche modello di questo sistema di assiomi. Se gli assiomi sono non-contraddittori, allora esistono molti universi

isomorfi; in alcuni di questi una data formula, ad esempio IC, può essere vera, in un altro falsa. In questo caso è indipendente dagli assiomi.

Una descrizione logica piú rigorosa deve essere formalizzata e il concetto di modello, in essenza, si riferisce a un'applicazione di due teorie formali l'una nell'altra. Si riporta una variante semiformalizzata, nella quale i modelli della teoria formale in L_1 Set saranno costituiti nella teoria intuitiva degli insiemi.

Si comincia con la descrizione dell'universo che si può considerare come il modello standard della teoria di Zermelo-Fraenkel.

4.1. Universo di Neumann-Zermelo-Fraenkel V.

Primi stadi. L'universo V è costruito per induzione a partire dall'insieme vuoto, per successive applicazioni dell'operazione $\mathcal{P}$ (insieme delle parti). Precisamente: $V_0=\Phi$, $V_1=\mathcal{P}(\emptyset)=\{\emptyset\}$, $V_2=\mathcal{P}(V_1)=\{\emptyset, \{\emptyset\}\}$, ..., $V_{n+1}=\mathcal{P}(V_n)$, ... È facile vedere che $V_n\subset V_{n+1}$; lo stadio V_n è costituito da $2^{2^{\cdot^{\cdot^{2}}}}$ (n-1 volte) insiemi finiti, i cui elementi sono insiemi finiti, «messi insieme» a partire da $\emptyset$. Per uscire dai limiti degli insiemi finiti, si ponga $V_{\omega_0}=\bigcup_{n=0}^{\infty} V_n$; dopo ciò si può iterare questa costruzione e porre $V_{\omega_0+1}=\mathcal{P}(V_{\omega_0})$, $V_{\omega_0+2}=\mathcal{P}(V_{\omega_0+1})$, ..., $V_{2\omega_0}=\bigcup_{i=1}^{\infty} V_{\omega_0+i}$, ...

Gli indici coi quali si numerano gli «stadi» di V sono gli ordinali di Cantor: insiemi totalmente ordinati a meno di isomorfismi. Prima di completare la descrizione, occorre dare gli elementi della teoria degli ordinali.

Ordinali. Gli insiemi ordinati sono stati definiti nel § 3 dell'articolo «Applicazioni». Sia $(X, <)$ un insieme ordinato; esso è detto linearmente ordinato se per ogni $Y,Z\in X$, con $Y\neq Z$, si ha $Y<Z$ oppure $Z<Y$; si dice totalmente ordinato se, oltre a ciò, ogni sottoinsieme non vuoto di X ha un elemento minimo.

Si dice isomorfismo degli insiemi ordinati X, X' un'applicazione biiettiva, strettamente crescente, fra loro. Si dice segmento iniziale dell'insieme totalmente ordinato X ogni sottoinsieme della forma $\hat{Y}=\{Y' \mid Y'<Y\}$, dove $Y\in X$. I segmenti iniziali di X sono totalmente ordinati dalla relazione $\subset$ in X.

Il risultato fondamentale sugli insiemi totalmente ordinati afferma che se X e Y sono due di questi insiemi, allora o sono isomorfi, oppure uno di essi è isomorfo a un segmento iniziale dell'altro. Per il momento si chiami ordinale una classe di insiemi totalmente ordinati, a meno di isomorfismi. Ogni coppia di ordinali è legata da una delle relazioni $\alpha=\beta$, $\alpha<\beta$, $\beta<\alpha$ a seconda che i loro rappresentanti siano isomorfi come insiemi totalmente ordinati, oppure uno sia isomorfo a un segmento iniziale dell'altro.

La formalizzazione del concetto di ordinale nel linguaggio L_1 Set si può realizzare comodamente seguendo Neumann. Secondo Neumann si chiama ordinale un insieme X totalmente ordinato dalla relazione $\in$ fra i suoi elementi e transitivo, vale a dire tale che da $Z\in Y\in X$ segua $Z\in X$.

Nel paragrafo successivo verranno identificati gli ordinali finiti con i numeri naturali. La classe di tutti gli ordinali è totalmente ordinata. L'ordinale minimo nella classe di tutti gli ordinali maggiori di α si indica con $\alpha+1$ e si chiama successore di α. Nella definizione di Neumann $\alpha+1=\alpha\cup\{\alpha\}$. L'ordinale β si chiama *limite* se non esiste un ordinale α tale che $\alpha=\beta+1$. Il primo ordinale limite ω_0 è isomorfo a $\{0, 1, 2, 3, \ldots\}$. Se β è un limite, nella definizione di Neumann $\beta=\bigcup_{\alpha<\beta}\alpha$.

Tutti gli ordinali costituiscono una classe totalmente ordinata, ma non un insieme: altrimenti la loro unione sarebbe l'ordinale piú grande, ma il seguente sarebbe ancora piú grande [Burali-Forti 1896; Cantor 1895-97].

Gli ordinali sono utilizzati per tre scopi principali: per le dimostrazioni mediante induzione transfinita, per le costruzioni mediante recursione transfinita e per le misure di potenza degli insiemi (cfr. «Applicazioni», § 2).

L'induzione transfinita si basa sul seguente principio: *la classe degli ordinali che contiene $\emptyset$, che accanto a ogni ordinale contiene il successore e che accanto a ogni insieme di ordinali contiene la sua unione, coincide con la classe di tutti gli ordinali.* L'applicazione del principio di induzione transfinita in questa forma solitamente richiede una scelta preventiva di un ordinamento totale dei vari insiemi; la sua esistenza è stata dimostrata da Zermelo per mezzo dell'assioma di scelta. Nelle ricerche contemporanee, soprattutto di algebra, si preferisce di solito usare il cosiddetto lemma di Zorn, il quale fa a meno della necessità di questa scelta.

Per le misure di potenza s'introduce il concetto di *cardinale*: si chiama cosí un ordinale che non sia equipotente ad alcun ordinale inferiore. Tutti gli ordinali fino a ω_0 compreso sono anche cardinali; ma ω_0+1, ω_0+2 e molti ordinali successivi sono numerabili e per questo non sono cardinali. Ogni insieme è equipotente a un unico cardinale e questi, in tal modo, costituiscono una scala di potenze naturale.

Il primo cardinale maggiore di ω_0 si designa con ω_1; il cardinale della potenza del continuo si denota con 2^{ω_0}; per questo l'ipotesi del continuo si può formulare come uguaglianza $\omega_1=2^{\omega_0}$. La classe di tutti i cardinali infiniti è totalmente ordinata e per questo è naturale numerarla mediante gli ordinali; se ω_α è il cardinale relativo al numero α, allora l'espressione $\omega_{\alpha+1}=2^{\omega_\alpha}$ (potenza di $\mathcal{P}(\omega_\alpha)$) si dice ipotesi del continuo generalizzata. I problemi sulla verificabilità di questo tipo di relazioni nell'aritmetica degli ordinali e dei cardinali, i problemi relativi ai cosiddetti cardinali inaccessibili e altri, formano la cerchia dei problemi dell'infinito. Quelli a cui non è stata data una soluzione, come è noto oggi, sono formalmente indecidibili. Le loro correlazioni coi piú tradizionali problemi matematici dell'analisi, della topologia o della teoria dei numeri sono comprese molto male.

Conclusione della costruzione dell'universo. Per ogni ordinale β poniamo $V_\beta=\mathcal{P}(V_\alpha)$, se $\beta=\alpha+1$; $V_\beta=\bigcup_{\alpha<\beta}V_\alpha$ se β è limite, infine $V=\bigcup_{\alpha}V_\alpha$. Questo è l'universo, costruito per recursione transfinita.

In particolare contiene tutti gli ordinali di Neumann: si può mostrare che $\alpha \in V_{\alpha+1}$.

La formalizzazione della costruzione descritta sopra consiste di due passi fondamentali: la costruzione della formula di L_1 Set designante «Per ogni insieme x esiste un ordinale α con $x \in V_\alpha$» e la deduzione di questa formula dagli assiomi di Zermelo-Fraenkel. I fondamenti logici sono tali da considerare che V sia un modello di questi assiomi. Di conseguenza alcune considerazioni intuitive mostrano che V è chiuso rispetto alle abituali operazioni sugli insiemi: differenza, unione, intersezione, formazione di $\mathcal{P}(X)$ e di $\bigcup_{Y \in X} Y$, formazione di $\{f(X) \mid X \in Y\}$, dove f è una funzione, formazione dei prodotti diretti, delle relazioni, delle applicazioni e delle immagini, ecc.

4.2. Funzione di verità e modelli.

Sia ora $M \subseteq V$ una classe di insiemi, sia A l'alfabeto di L_1 Set, $A_0 \subset A$ un insieme di simboli di variabile. Si dice M-specializzazione una qualunque applicazione ξ di A_0 in M, vale a dire la scelta, per ogni variabile $x \in A_0$, di uno dei suoi valori x^ξ, il quale è un insieme di M. Per induzione sulla lunghezza della formula P del linguaggio L_1 Set si definisce il valore della funzione di verità $|P|_M(\xi) = 1$ (vero) oppure $|P|_M(\xi) = 0$ (falso). Precisamente si ponga $|x \in y|_M(\xi) = 1$ se $x^\xi \in y^\xi$; 0 se $x^\xi \notin y^\xi$; analogamente per le altre formule atomiche. Quindi si ponga $|\neg P|_M(\xi) = 1 - |P|_M(\xi)$, $|P \wedge Q|_M(\xi) = |P|_M(\xi) \cdot |Q|_M(\xi)$, ecc., secondo le abituali regole della logica; infine si ponga $|\exists x P|_M(\xi) = \max |P|_M(\xi')$, $|\forall x P|_M(\xi) = \min |P|_M(\xi')$, dove ξ' percorre tutte le M-specializzazioni e si distingue da ξ solo per il valore della variabile x.

La formula P si dice M-vera se $|P|_M(\xi) = 1$ per ogni ξ, e M-falsa se $|P|_M(\xi) = 0$ per ogni ξ. Ogni formula chiusa è M-vera oppure M-falsa: il valore della sua funzione di verità non dipende da ξ.

La classe M si dice modello dell'insieme di formule $\mathfrak{S}$ (ad esempio di un sistema di assiomi), se tutte le formule a partire da $\mathfrak{S}$ sono M-vere. Solitamente si considerano classi transitive, per le quali da $Y \in X \in M$ segue $Y \in M$.

La formalizzazione del concetto di modello sostituisce il concetto di classe M col concetto di formula $Q(x)$ con una variabile libera, la quale esprime M in L_1 Set, e la verità alla deducibilità dagli assiomi.

Piú in particolare, si definisce la Q-relativizzazione P_Q della formula P del linguaggio L_1 Set per induzione sulla sua lunghezza. Per le formule atomiche: $(x \in y)_Q$ è $Q(x) \wedge Q(y) \to x \in y$, $(x = y)_Q$ è $Q(x) \wedge Q(y) \to x = y$. Quindi $(\neg P)_Q$ è $\neg(P_Q)$; $(P_1 * P_2)_Q$ è $(P_1)_Q * (P_2)_Q$ per ogni connettivo $*$; $(\forall x P)_Q$ è $\forall x (Q(x) \to P)$ e $(\exists x P)_Q$ è $\exists x (Q(x) \wedge P)$. Una formula si dice transitiva se dagli assiomi è deducibile la formula $Q(x) \wedge y \in x \to Q(y)$.

La formula Q definisce un modello interno del sistema di assiomi se tutte le Q-relativizzazioni degli assiomi sono deducibili da questi. Questo modello è transitivo se Q è transitiva.

Verranno ora descritti alcuni modelli che risolvono dei problemi dell'infinito.

4.3. Costruzione dell'universo di Gödel L.

Gödel definí una sottoclasse $L \subseteq V$ molto interessante: l'universo degli insiemi costruibili. La sua piú semplice definizione è la seguente: si tratta della piú piccola sottoclasse di V contenente tutti gli ordinali e che sia un modello per gli assiomi di Zermelo-Fraenkel. Tuttavia questo fatto si ottiene assai piú tardi e come conseguenza di un'altra definizione intuitiva, la quale si usa anche nelle dimostrazioni. Per definirla, vanno introdotte prima le operazioni di Gödel sui sottoinsiemi. Si tratta di tre operazioni binarie (formazione della coppia, differenza e prodotto diretto), due operazioni unarie (formazione dell'insieme dei primi elementi di tutte le coppie ordinate contenute in un dato insieme, e formazione del grafico di una relazione di inclusione nel quadrato diretto di un dato insieme) e sei operazioni unarie (formazione dell'insieme delle terne ordinate, ottenute con una permutazione dalle terne ordinate contenute in un dato insieme).

Piú avanti s'indicherà con $\bar{X}$ la chiusura di qualunque insieme X: è il piú piccolo insieme contenente tutti gli elementi di X in maniera chiusa rispetto alle operazioni di Gödel.

Ora la definizione effettiva di universo L è questa: $L = \bigcup L_\alpha$ (unione rispetto agli ordinali α), dove $L_0 = \varnothing$, $L_{\alpha+1} = \mathcal{P}(L_\alpha) \cap (L_\alpha \cup \{L_\alpha\})^-$, $L_\beta = \bigcup_{\alpha<\beta} L_\alpha$, se β è definito. In altri termini ad ogni passo non si aggiungono tutti i sottoinsiemi dello stadio già costruito, ma solo quelli che si possono ottenere dagli insiemi a disposizione mediante le operazioni di Gödel.

I primi stadi di L e di V coincidono: $L_\alpha = V_\alpha$ per $\alpha \leq \omega_0$; tuttavia a partire da $\alpha = \omega_0 + 1$ l'inclusione $L_\alpha \subset V_\alpha$ diventa propria. In effetti si può dimostrare che la potenza di L_α è uguale alla potenza α, già quando la potenza di V_{ω_0+1} è uguale a 2^{ω_0}. Da un punto di vista logico ciò non esclude la possibilità che $\forall \alpha \exists \beta (V_\alpha \subset L_\beta)$, vale a dire $L = V$.

La proposizione $L = V$, piú precisamente la sua traduzione formale nel linguaggio L_1 Set, si chiama «assioma di costruibilità». Ha uno status molto interessante. Gödel ha dimostrato che non contraddice gli assiomi di Zermelo-Fraenkel se questi stessi sono non-contraddittori, tuttavia sottolineò come non esista alcun fondamento per assumerlo «vero». Evidentemente egli era intuitivamente convinto che in V si trovi una grande quantità di insiemi non costruibili, di modo che l'assioma di costruibilità sarebbe falso.

Gödel dimostrò che dagli assiomi ZF1-ZF9 e dall'assioma $L = V$ è deducibile l'assioma di scelta ZF10. Intuitivamente ciò corrisponde alla possibilità di costruire in maniera ricorsiva una «funzione universale di scelta» per ogni classe L, la quale ponga in corrispondenza ad ogni insieme costruibile non vuoto un suo elemento e che sia essa stessa costruibile.

Infine Gödel dimostrò che dagli assiomi ZF e dall'assioma $L = V$ si deduce l'ipotesi del continuo generalizzata, che in tal modo non contraddice gli assiomi di Zermelo-Fraenkel, se questi stessi sono non-contraddittori. La parte

«intuitiva» fondamentale di questa dimostrazione stabilisce che la potenza di $\mathcal{P}(\omega_\alpha)\cap L$ è $\omega_{\alpha+1}$, vale a dire i sottoinsiemi costruibili di un cardinale sono tanti quanto il cardinale successivo. Tuttavia se è «proprio» $L\neq V$, vi possono essere molti piú sottoinsiemi.

L'assunzione dell'assioma $L=V$ si può sostituire intuitivamente col calcolo delle funzioni di L-verità per una serie di formule del linguaggio L_1 Set. Ciò permette di fare la dimostrazione di Gödel in maniera meno sintattica, ma rimane comunque una lunga e difficile componente sintattica, connessa all'effetto di «variazione di significato» delle formule del linguaggio formale nei modelli non standard. Si dimostrerà questo effetto sul cosiddetto «paradosso di Skolem», il quale è tecnicamente piú semplice e fu scoperto prima del modello di Gödel.

4.4. Modelli numerabili e paradosso di Skolem.

Sia $M\subseteq V$ una sottoclasse di V. Stabilendo che M è un modello degli assiomi, si calcoli la funzione $|P|_M(\xi)$ per ogni assioma P, sapendo che $|P|_V(\xi)=1$. Il concetto tecnico fondamentale risulta qui quello di formula assoluta: P si dice M-assoluta se per ogni M-specializzazione ξ si ha $|P|_M(\xi)=$ $=|P|_V(\xi)$. È chiaro che se la formula P è V-vera e M-assoluta, allora è anche M-vera, mentre quello di essere assoluta è un concetto «piú sintattico» e per questo risulta piú comodo verificarlo separatamente. Comunque ciò non è semplice come $L=V$ se la formula non è scritta chiaramente, ma data soltanto con una complessa costruzione intuitiva. In ciò si trova la radice delle difficoltà sintattiche della dimostrazione di Gödel.

Il metodo universale per spezzare l'assolutezza è il seguente: se la formula $Q(x)$ contiene una variabile libera, allora $\exists x Q(x)$ può essere V-vera, in quanto in V si trova un insieme con la proprietà Q, ma non M-vera, poiché in M un simile insieme non si trova.

Per questo, per conservare l'assolutezza, ad esempio, di un insieme numerabile di formule di L_1 Set, può risultare necessario aggiungere un insieme per ogni formula vera $\exists x Q(x)$ alla classe scelta originariamente $M_0\subseteq V$. Dopo ciò possono variare i valori delle funzioni di verità e di altre formule, ma l'iterazione di una quantità numerabile di simili passi è già sufficiente a garantire l'assolutezza della classe ampliata $M\supseteq M_0$.

In questo modo è possibile costruire un modello numerabile $M\subseteq V$ degli assiomi di Zermelo-Fraenkel, a partire almeno da $M_0=\{\emptyset\}$. Per la verità questo può non essere transitivo: alcuni insiemi di M conterranno elementi non appartenenti a M. Ma, come ha dimostrato Mostowski, con un procedimento di eliminazione degli elementi «estranei» si può costruire un M, isomorfo rispetto alla relazione $\in$, che sia una sottoclasse transitiva di V, il quale, in tale modo, è un modello transitivo numerabile della teoria degli insiemi di Zermelo-Fraenkel. Si capisce che i principî descritti brevemente qui sono molto generali, e *mutatis mutandis*, si riferiscono ai modelli insiemistici di altri linguaggi formali numerabili.

Verranno ora presentate alcune implicazioni dell'esistenza di modelli numerabili della teoria degli insiemi.

Nel modello M c'è un insieme infinito x_0, come segue dalla M-verità dell'assioma di infinità. In M c'è anche «l'insieme di tutte le M-parti di x_0» $\mathcal{P}(x_0)_M$ in forza della verità dell'assioma di potenza. Sia x_0 sia $\mathcal{P}(x_0)_M$ non sono piú che numerabili, giacché il modello M è transitivo e numerabile. Per questo deve esistere un'applicazione di x_0 su $\mathcal{P}(x_0)_M$. Tuttavia anche il teorema di Cantor è deducibile dagli assiomi, vale a dire anche la formula «non esiste un'applicazione di x_0 su $\mathcal{P}(x_0)$» è M-vera. Questi fatti non si contraddicono l'un l'altro, né contraddicono il senso delle corrispondenti proposizioni in V, poiché $\mathcal{P}(x_0)_M=\mathcal{P}(x_0)\cap M$, e ciò rende $\mathcal{P}(x_0)_M$ non piú che numerabile, e poiché la M-verità del teorema di Cantor significa, «dal punto di vista di V», soltanto assenza di quelle applicazioni di x_0 su $\mathcal{P}(x_0)_M$ che sono a loro volta elementi di M.

Per motivi storici si è preso a chiamare «paradosso di Skolem» tutti questi effetti di variazione di significato degli enunciati nei modelli, in particolare in quelli numerabili. Alcuni autori, fra cui lo stesso Skolem, pronti a lavorare con infinità numerabili, ma non superiori, consideravano questi effetti come una manifestazione di relatività dei concetti insiemistici, in particolare del continuo.

Gli specialisti di analisi e di topologia, abituati a lavorare col continuo reale come con un qualsiasi formalismo della realtà, vedono nel paradosso di Skolem soltanto una manifestazione di inadeguatezza dei formalismi, in quanto metodi per imitare le considerazioni intuitive. Questa convinzione è sostenuta soltanto dall'esame di effetti analoghi nei modelli non standard del linguaggio dell'aritmetica, in cui l'intuitiva interpretazione standard appare alla maggior parte dei matematici assolutamente attendibile e univoca.

In ogni caso, la costruzione di speciali modelli numerabili è un metodo potente per dimostrare l'indipendenza di diversi problemi di teoria degli insiemi. Il metodo generale di costruzione di questi modelli, il *forcing*, fu scoperto da Cohen. Lo si descrive però in un'altra variante, secondo Scott e Solovay, i quali utilizzano considerazioni intuitive molto interessanti.

4.5. Universo sopra un'algebra di Boole.

Algebre di Boole. Sia S un insieme qualsiasi, $B=\mathcal{P}(S)$ l'insieme delle sue parti. In B si distinguono gli elementi $\emptyset=0$ e $S=1$; inoltre in B vi sono due operazioni binarie $\cup$ e $\cap$ e una operazione unaria, il passaggio al complemento. Le operazioni $\cup$, $\cap$, $'$ sono legate da alcune identità, ad esempio $(a\cap b)'=a'\cup b'$ oppure $a\cap(b\cup c)=(a\cap b)\cup(a\cap c)$. L'insieme $\mathcal{P}(S)$ con queste operazioni e gli elementi distinti 0, 1 è un'algebra di Boole.

Piú in generale un insieme B con le operazioni $\wedge$, $\vee$, $'$ e con gli elementi distinti 0, 1 si dice algebra di Boole se in esso è verificata ogni identità di un elenco finito (descritto nell'articolo «Dualità», § 1, di questa stessa *Enciclopedia*).

Sia ad esempio S uno spazio con misura μ (cfr. l'articolo «Applicazioni», § 3),

$\mu(S)=1$, $B\subset\mathcal{P}(S)$ sia l'algebra degli insiemi misurabili, $B=\bar{B}$ mod o l'algebra delle classi di insiemi misurabili rispetto alla relazione di equivalenza «misura della differenza simmetrica uguale a zero». Allora B costituisce un'algebra di Boole; queste algebre svolgono un notevole ruolo nei modelli matematici della teoria delle probabilità. Ad esempio, rappresentazioni intuitive delle «grandezze casuali» sono ottenute col concetto matematico di funzione f sullo spazio S; la probabilità che la misura del valore della funzione f sia compresa fra a e b si ottiene dal numero $\mu(\{x\in S \mid a\leq f(x)\leq b\})$ (si suppone che tutte queste misure esistano, vale a dire che f sia una funzione misurabile).

In analogia con questo esempio è comodo figurarsi che gli elementi dell'universo, descritto piú avanti, sull'algebra di Boole B, siano «insiemi casuali».

Universo V^B e «verità booleana». Anche questo universo, come V, si costruisce per stadi, numerati dagli ordinali: $V^B=\bigcup_\alpha V^B_\alpha$. Per ogni coppia di elementi $X,Y\in V^B$ si definiscono induttivamente le «funzioni di verità a valori booleani» $\|X\in Y\|\in B$ e $\|X=Y\|\in B$, che occorre rappresentarsi come «probabilità» di appartenenza di X a Y o di uguaglianza fra X e Y (nel consueto linguaggio delle probabilità le *misure* degli elementi dell'algebra B, e non gli elementi stessi, si dicono probabilità). Prima di descrivere esplicitamente la costruzione, dipendente solo da B, verrà indicato come la si utilizza nelle dimostrazioni di indipendenza.

Come prima, sia A_0 un insieme di simboli di variabile del linguaggio L_1 Set; si chiami V^B-specializzazione ξ un'applicazione di A_0 in V^B. Per ogni ξ e per ogni formula P del linguaggio L_1 Set è possibile definire il «valore di verità booleano» $\|P\|(\xi)$ per induzione sulla lunghezza di P. Per le formule atomiche la definizione è esposta nella costruzione dell'universo V^B: si tratta delle corrispondenti «probabilità». Quindi $\|\neg P\|=\|P\|'$, $\|P\wedge Q\|=\|P\|\wedge\|Q\|$, ecc. Questa associazione di valori di verità booleani agli enunciati è molto vicina alle idee originarie di Boole, uno dei fondatori della moderna logica formale.

Si dice che la formula P è B-vera se $\|P\|(\xi)=1$ per ogni ξ, e B-falsa se $\|P\|(\xi)=0$ per ogni ξ. Dal punto di vista delle probabilità una formula è B-vera se l'enunciato che le corrisponde risulta in ogni caso «vero con probabilità 1».

Tutti gli assiomi logici del linguaggio L_1 Set sono B-veri per ragioni abbastanza banali, già scoperte da Boole. I principî esposti nella costruzione di V^B permettono di stabilire anche la B-verità di tutti gli assiomi di Zermelo-Fraenkel. Infine la B-verità si conserva per deduzioni a partire da formule B-vere.

Per questo nella dimostrazione di non-deducibilità di una qualunque formula, ad esempio IC, a partire dagli assiomi, è sufficiente scegliere un'algebra B tale che $\|\mathrm{IC}\|=0$. Anche questo si realizza col metodo di Cohen-Solovay che verrà descritto piú avanti.

Costruzione di V^B. Si procede per recursione transfinita su α. Posto $V^B_0=\varnothing$, si supponga che per l'ordinale α sia già stato definito l'insieme V^B_α; per ciascuno dei suoi elementi $X\in V^B_\alpha$ sia definito un sottoinsieme $D(X)\subseteq V^B_\alpha$; per ogni coppia di elementi $X,Y\in V^B_\alpha$ siano definite le «probabilità» $\|X\in Y\|\in B$, $\|X=Y\|\in B$. Si supponga inoltre che questi dati soddisfino le seguenti condizioni: se $\beta_1\leq\beta_2\leq\alpha$, allora $V^B_{\beta_1}\subseteq V^B_{\beta_2}$; se $\beta<\alpha$, $X\in V^B_{\alpha+1}\setminus V^B_\beta$, allora $D(X)=V^B_\beta$. Infine si supponga che

$$\|X\in Y\|=\bigvee_{Z\in D(Y)}(\|X=Z\|\wedge\|Z\in Y\|)$$

$$\|X=Y\|=(\bigwedge_{Z\in D(X)}\|Z\in X\|'\vee\|Z\in Y\|)\wedge(\bigwedge_{Z\in D(Y)}\|Z\in Y\|'\vee\|Z\in X\|).$$

È ora possibile definire tutti i dati necessari allo stadio $V^B_{\alpha+1}=V^B_\alpha\cup V^{B*}_{\alpha+1}$. Precisamente $V^{B*}_{\alpha+1}$ sarà costituito da tutte le funzioni possibili Z, con dominio di definizione V^B_α e a valori in B, le quali soddisfano alla condizione di estensionalità: $\|X=Y\|\wedge Z(X)=\|X=Y\|\wedge Z(Y)$ per ogni $X,Y\in V^B_\alpha$. Oltre a ciò per $Z\in V^{B*}_{\alpha+1}$ si ponga $D(Z)=V^B_\alpha$ e per le precedenti Z la definizione di $D(Z)$ non cambia.

Infine per gli ordinali limite β si ponga $V^B_\beta=\bigcup_{\alpha<\beta}V^B_\alpha$.

Un ulteriore lavoro consiste nella verifica induttiva del fatto che le identità per $\|X\in Y\|$ e $\|X=Y\|$ si conservano allo stadio $\alpha+1$; ciò conclude di fatto la costruzione dell'universo $V^B=\bigcup_\alpha V^B_\alpha$.

Si osservi che per formalizzare questa costruzione occorre considerare che $B\in V$. Se $B=(0,1)$ l'universo ottenuto sarà essenzialmente isomorfo a V, sebbene non coincida esattamente con esso.

Verifica della «verità» degli assiomi. Nella costruzione dell'universo V^B è stata posta direttamente la condizione di B-verità dell'assioma di estensionalità. Le tre identità scritte sopra per la funzione di verità garantiscono la B-verità delle conseguenze di questo assioma: $x\in y\leftrightarrow\exists z(x=z\wedge z\in y)$; $x=y\leftrightarrow(z\in x\rightarrow z\in y)\wedge(z\in y\rightarrow z\in x)$; $(x=y)\wedge(x\in z)\leftrightarrow(x=y)\wedge(y\in z)$ rispettivamente, dalle quali si deduce la sua B-verità.

La possibilità di garantire ciò è ben lungi dall'essere evidente: cosa significa che «gli insiemi coincidono quando sono costituiti dagli stessi elementi», quando sia «insiemi coincidenti» sia «essere elementi» sono enunciati che possono essere veri soltanto con una certa probabilità? Qui i formalismi, sia il linguaggio stesso sia le sue interpretazioni a valori booleani, appaiono invarianti rispetto alla possibilità di «scrivere le idee sotto forma di calcoli».

Tuttavia, dopo la verifica dell'assioma di estensionalità, quella dei rimanenti assiomi, anche se tecnicamente difficile, risulta un compito abbastanza diretto.

Scelta dell'algebra B che «confuta» l'ipotesi del continuo. È stato spiegato prima che la L-verità dell'ipotesi del continuo nella dimostrazione di Gödel era garantita dall'assunzione di una «minore» quantità generale di insiemi, in

particolare dei sottoinsiemi di ogni cardinale, ottenibile considerando soltanto i sottoinsiemi costruibili.

L'introduzione dell'algebra B produce l'effetto contrario: essa «aumenta» la quantità di insiemi, per tener conto della possibilità di variare le «probabilità» con cui un insieme è costituito dai suoi elementi. Questo approccio particolare permette anche di costruire, ad esempio, ω_1 «B-sottoinsiemi» di cardinalità ω_0, scegliendo l'algebra B di potenza abbastanza grande. Tuttavia, come anche nella dimostrazione di Gödel o del paradosso di Skolem, con le B-interpretazioni si ottiene una variazione di significato di tutte le proposizioni, non necessariamente in senso vantaggioso. Per questo il controllo di tutti questi effetti richiede un lavoro ancora maggiore, i cui particolari verranno qui evitati.

Si osservi per concludere che sono note anche diverse varianti di universo del tipo V^B, nelle quali il ruolo di «base» non è svolto da algebre di Boole, ma da insiemi ordinati o in generale da «topos» di Grothendieck. Alcune di queste varianti si possono ritenere come la realizzazione delle idee della matematica intuizionista coi metodi della matematica classica.

5. *Critica e alternative.*

Per quanto siano cambiati storicamente il contenuto e la forma dei testi matematici nel corso di due millenni, il loro carattere deduttivo rimane immutabile: «Dopo i Greci, chi dice matematica dice dimostrazione» [Bourbaki 1954, p. 1]. Per questo ogni critica e ogni controcritica sui fondamenti, fra cui la teoria degli insiemi, procedono lungo due linee: la validità dei concetti e delle premesse iniziali e l'aspettativa delle deduzioni (compresi i metodi per ottenerle).

Le argomentazioni riportate comportano invariabilmente un carattere non matematico. La validità della scelta dei concetti iniziali e delle regole d'analisi è verificata per confronto con una «realtà», vale a dire appellandosi al significato di un testo matematico. Perfino per la filosofia formalista della matematica è cosí, giacché la scelta dei formalismi, in particolare situati entro la loro logica, e la loro metateoria sono fondati su considerazioni di contenuto.

Tuttavia la categoria del significato del testo matematico, di per sé, è un'astrazione d'ordine superiore. A uno studio piú immediato si presentano i procedimenti di interpretazione del testo, che si assumeranno quale materiale originario. In questi procedimenti si distinguono alcuni aspetti, che a loro volta si manifestano in diversi atti individuali di interpretazione. Li si descriverà e si proseguirà mostrando come l'assolutizzazione di uno qualsiasi di questi aspetti definisca ciascuna delle posizioni filosofiche adottate sui problemi dei fondamenti.

L'uomo impara dall'infanzia l'interpretazione dei testi correnti nella lingua natale, parallelamente al possesso della stessa lingua, quando si trova in un ambiente linguistico omogeneo. In questo modo nella coscienza individuale si forma un nucleo rapido e univoco di testi comprensibili. Questa indivisibilità del processo di comprensione dei testi quotidiani permette di assegnarle la qualità di significato determinato. Naturalmente quando questo processo di comprensione diventa l'oggetto di studio di psicologi, linguisti, psichiatri, rivela una complessità enorme e una organizzazione gerarchica. Ma per gli scopi di questo articolo basta considerarlo come unitario e considerare il significato prodotto dall'interpretazione come un'autentica astrazione.

Il problema si presenta assai piú complesso con i linguaggi della scienza. Nella formazione di un abituale concetto delle scienze naturali, di un sistema di concetti o di una teoria, e anche nel loro studio, i testi corrispondenti presentano anche le istruzioni per la loro interpretazione. Queste istruzioni possono costituire una parte del testo ben delimitata, ma possono anche essere mescolate in costruzioni linguistiche, parallele a costruzioni linguistiche di testi già interpretati. Come se si riuscisse a capire l'espressione «L'elettrone è passato oltre la fenditura A (e non la fenditura B)» nei manuali di meccanica quantistica, avendo come «istruzione» per l'interpretazione la sua sintassi, parallela alla sintassi della frase, comprensibile in maniera univoca, «Il sasso è passato oltre la fenditura A e non la B». Questa «istruzione» induce in errore e numerose spiegazioni seguenti sono dedicate all'introduzione della frase, ben formata nel linguaggio della meccanica quantistica, «Ampiezza di passaggio di un elettrone oltre la fenditura A», e alle istruzioni per interpretarla. Proprio alla fine di questa gerarchia di istruzioni si trovano solitamente le descrizioni dei procedimenti operativi delle scienze sperimentali: misurazioni, osservazioni, ecc. L'interpretazione di un testo comprende una, piú o meno completa, comprensione di questa gerarchia.

L'originalità dei testi matematici consiste solitamente nell'assenza di questi procedimenti operativi, ai quali il contenuto del testo si riferisce, magari in ultima analisi. Ciò nondimeno in determinate situazioni questo aspetto «realistico» svolge un ruolo fondamentale nell'interpretazione dei testi matematici.

Inoltre, interpretando un testo matematico, il lettore utilizza forme visive, cinematiche e altre, piú o meno determinate, che sorgono nella sua coscienza. L'aspetto corrispondente del significato lo si chiama «interno».

Finalmente il testo interpretato nelle sue relazioni con altri testi, reali o potenziali, vale a dire il costrutto delle sue traduzioni in altri sottolinguaggi, l'esplicitazione di definizioni, le formalizzazioni, ecc., rivolgono il matematico verso l'aspetto «esterno» del significato.

Si consideri come semplice esempio l'interpretazione del concetto «due» nei contesti «prendi due mele», «dualismo filosofico» e «la sesta cifra dopo la virgola nello sviluppo decimale di π è due». Nel primo è piú essenziale l'aspetto realistico del significato. Nel secondo, quello interno, giacché la rappresentazione di «due» deve applicarsi a un sistema concettuale tanto complesso quanto «le essenze originali, giacenti a priori a fondamento del mondo, all'interno dei sistemi filosofici di una determinata classe». Infine nel terzo contesto domina il significato esterno, poiché la comprensione di questo testo fa appello alla ricostruzione di un certo metodo di calcolo approssimato, di un metodo per ottenere le cifre decimali della corrispondente approssimazione, di un metodo di valutazione della precisione, ecc.

Ecco ora una descrizione piú dettagliata di questi aspetti del significato e della loro influenza sulla scelta e sulla valutazione dei diversi sistemi di fondamenti della matematica.

5.1. Aspetto realistico della semantica.

Non v'è dubbio che svolga un ruolo fondamentale nella matematica «applicata», intesa in senso lato. Rientrano qui, in modo naturale, i calcoli ingegneristici ed economici. Si riferisce anche all'apparato della fisica teorica, seppure qui sia nettamente ingigantito il ruolo e la complessità dei procedimenti di esplicazione del significato fisico dei calcoli. Per questo il rigore matematico-formale dei calcoli può passare, in fisica, in secondo piano, quando le regole d'interpretazione costituiscono un fattore formativo piú forte per la teoria. Una situazione come quella delle procedure di rinormalizzazione in elettrodinamica quantistica, che per la matematica appaiono come regole introdotte ad hoc nel testo formale, hanno un significato e una giustificazione definitiva nella coincidenza dei calcoli correttivi radiazionali con le misure sperimentali. La fisica costituisce il piú importante modello di pluralità d'interpretazioni realistiche della stessa teoria matematica: l'equazione delle onde può descrivere la vibrazione di una corda e la diffusione degli elettroni; un integrale può corrispondere a un volume, a una carica, a un percorso o a una probabilità; gli esempi si possono moltiplicare senza fine. L'accentuazione di questa idea conduce a concezioni tanto diverse quanto il «convenzionalismo» di Poincaré, il «relativismo insiemistico» di Skolem o l'interpretazione insiemistica della logica intuizionista.

Per la matematica «pura» l'aspetto realistico del significato può essere dominante, quando il testo è interpretato come la descrizione di un procedimento algoritmico, condotto sul testo stesso. Sono di questo tipo gli algoritmi elementari con le frazioni decimali o le regole matematiche di sostituzione di un termine in una formula. Infine il significato del programma per un calcolatore è quella successione di stati, prodotta dal calcolatore quando opera seguendo quel programma. Con un po' d'esagerazione si può dire che il lavoro di un calcolatore è anche uno dei procedimenti d'interpretazione del programma.

La matematica egiziana e babilonese era quasi completamente «realistica» e tale è anche la semantica dei tradizionali manuali scolastici per le classi inferiori.

Occorre osservare che, nel nostro senso, la parola 'realistico' va interpretata in modo esteso. Perfino quando la sua interpretazione si effettua in termini di procedimenti empirici, vengono sottintese in realtà certe astrazioni, spesso di livello molto elevato. Queste astrazioni possono riferirsi non solo alle scienze naturali, ma anche alla filosofia: il presupposto che esista un mondo «neoplatonico» degli insiemi permette di annoverare nella filosofia «realistica» la classica matematica insiemistica. La negazione di questo mondo, in quanto non reale, e la sua sostituzione col piú reale mondo dei testi e dei procedimenti algebrici su di essi, caratterizza «l'altro realismo», quello della matematica costruttiva. L'orientamento sulla realtà del mondo fisico e l'interpretazione della matematica in termini di scienze naturali predetermina la filosofia «applicata» della matematica, a stento formulata sistematicamente ma creata quotidianamente dai fisici.

5.2. Aspetto interno della semantica.

Come si è detto, si manifesta in ogni atto individuale d'interpretazione e anche nel compito di esplicare il «significato intuitivo del testo», dipendente dalla coscienza individuale.

Ogni persona ha accesso immediato solo alla propria coscienza. L'autoosservazione e i metodi linguistici permettono di formare un modello esterno dei propri atti cognitivi, anche se del tutto inadeguato e incompleto. Inoltre la componente inconscia fondamentale esce completamente dalla descrizione, oppure si deforma a seguito dei tentativi di esteriorizzarla.

Per questo le descrizioni esistenti del significato intuitivo dei testi si riferiscono abitualmente a concetti separati e ad atti cognitivi elementari. Oltre ai filosofi e agli psicologi, hanno formato dei modelli notevoli di tali descrizioni Pascal, Poincaré, Hadamard. La struttura anatomica e funzionale del sistema nervoso centrale, comune a tutti gli uomini, da un lato, e dall'altro, le comuni esperienze e consuetudini linguistiche, permettono di isolare una totalità di unità espressive elementari per «il linguaggio dei significati intuitivi». Fra queste possono rientrare termini paramatematici come 'quantità', 'vicinanza', a un livello piú alto 'relazione', 'continuo/discreto', a un livello ancora superiore 'algoritmo'. In qualità di unità espressive possono agire anche disegni schematici.

L'assegnazione dello status di termine di un sottosistema assiomatizzato della matematica porta il concetto intuitivo dal livello di «semantica interna» al livello di «semantica esterna». Inoltre la precisione del significato procede per mezzo dell'eliminazione di possibilità, in potenza racchiuse nel modello intuitivo. Cosí il concetto intuitivo di «grandezza attuale infinitamente piccola» fu completamente eliminato nei fondamenti classici dell'analisi e comparve nei concetti precisi di «elemento nilpotente» della geometria algebrica e della geometria differenziale dei nostri tempi, e di «infinitesimo» nei modelli non-standard dei numeri reali da parte dei logici. Conservando alcuni tratti dell'archetipo intuitivo, queste interpretazioni «esterne» delle grandezze infinitamente piccole non coincidono e portano in direzioni diverse.

La totalità delle unità del significato interno è individualmente e storicamente variabile, aperta e debolmente strutturata. La conquista di nuovo materiale porta alla formazione di nuove unità.

Ad esempio la mutua interazione fra le particelle elementari e col campo, può essere descritta in fisica nel linguaggio dei «grafi di Feynmann». Ciascuno di questi grafi «è» una configurazione di punti di vario genere, uniti da linee-lati di vario genere. Interpretando «realisticamente» un dato grafo, un fisico vede in parte della sua struttura le linee universali delle particelle elementari. Un'al-

tra parte della struttura, che non ammette questa interpretazione, sarà paragonata a particelle «virtuali». Cosí sorge una nuova, comoda, unità di significato intuitivo, rispondente ai fatti osservati, ma intermedia rispetto alle unità espressive del linguaggio accettato. In modo analogo sono sorti i quark, corrispondenti alle rappresentazioni irriducibili di un certo gruppo. Il problema dell'«esistenza reale» di particelle virtuali o di quark necessita di una sottile interpretazione, prima che si tenti di risolverlo.

Si osservi infine che l'interpretazione «esterna» del grafo di Feynmann presuppone, ad esempio, la capacità di scrivere un termine che gli corrisponde nella teoria delle perturbazioni, e l'interpretazione «esterna» dei quark, la capacità di decomporre un prodotto tensoriale di due rappresentazioni nella somma delle loro componenti irriducibili.

Ogni concezione coerente dei fondamenti della matematica tende a limitare al minimo necessario la base di unità di significato intuitivo che viene adottata. Si producono con ciò una scelta di unità espressive accettabili e una limitazione nella dimensione del loro contenuto intuitivo, dopo di che le rimanenti unità da assumere sono strutturate per mezzo di quelle elementari.

L'accettazione del linguaggio della teoria degli insiemi, preparato dall'«aritmetizzazione del continuo», ha in effetti permesso di diminuire notevolmente il campo delle unità elementari del significato «interno» o, secondo Bourbaki, «delle varie intuizioni», necessario alla matematica, senza diminuire il volume della matematica classica. Dopo aver imparato l'interpretazione intuitiva dei concetti e delle costruzioni insiemistiche, il matematico poté cessare di rivolgersi all'intuizione come all'unica fonte di molti altri concetti. Ad esempio la «curva» diventava complessa, ma rientrava in modo inequivocabile nei costrutti insiemistici.

Il successivo approfondimento di questo concetto poté scoprire proprietà «intuitivamente evidenti» in precedenza, o anche del tutto nuove, come l'esistenza della curva di Peano, che riempie un quadrato, la quale appariva nettamente contraria all'intuizione (Poincaré). La teoria degli insiemi ha salvato la maggior parte dei matematici dall'anarchia della diversità delle intuizioni individuali, lasciando per la verità questo fardello a una minoranza, che tenta di ottenere una spiegazione intuitiva e un accordo sul significato interno dell'unico fondamentale concetto di insieme, che inizialmente era cosí intuitivo.

Si considerino da questo punto di vista alcune alternative al linguaggio della teoria degli insiemi. La prima è il linguaggio duale delle «proprietà», dei «predicati», ecc. Il suo enunciato elementare, «avere una proprietà», sostituisce l'enunciato «appartenere a un insieme». La semantica del concetto di proprietà, assunto come iniziale, non è intuitivamente obbligata ad essere completamente equivalente alla semantica del concetto «insieme di elementi con una data proprietà», almeno per il fatto che la determinatezza del concetto «essere rosso» non richiede affatto la determinatezza «dell'insieme degli oggetti rossi», che tenga conto almeno della presenza di una scala di sfumature di colore. Oltre a ciò, nella matematica precedente a Cantor, il linguaggio delle proprietà era usato proprio per sostituire l'infinità attuale, diventata un tabú. In

effetti, intuitivamente, la «verifica del fatto che un oggetto possiede una data proprietà» si distingue nettamente dalla «raccolta di tutti gli oggetti con quella data proprietà».

Ciò nondimeno, nei sistemi di Russell-Whitehead, di Quine e di altri autori, la formalizzazione del linguaggio delle proprietà nella «teoria dei tipi» era usata precisamente per fondare la teoria degli insiemi. Lo scopo pragmatico consisteva in una limitazione dell'universo e dei metodi linguistici, tale da rendere impossibile la presenza di paradossi del tipo di quello di Russell.

Per la teoria dei tipi la «realtà» è l'universo, costituito da individui (di tipo zero), loro proprietà (di primo tipo), proprietà delle loro proprietà (di secondo tipo) e cosí via senza fine. Il paradosso di Russell non può verificarsi poiché la proprietà «essere una proprietà, non posseduta da se stessa» non appartiene all'universo.

Tuttavia in questo universo non è comodo interpretare la matematica classica, già al livello del linguaggio, ed è quasi impossibile al livello degli assiomi di esistenza: l'assioma di «riducibilità» introdotto da Russell non risulta soddisfacente.

Varianti della teoria dei tipi sono connesse alla tesi «logicista», secondo la quale la matematica è riducibile alla logica, per motivi storici e intuitivi; questa tesi stessa possiede un contenuto appena percettibile.

Nella matematica intuizionista, il posto sia di «insieme» che di «proprietà» è occupato all'incirca dall'enunciato «legge di formazione». Tuttavia l'assunzione fondamentale dell'intuizionismo, relativa alla non-essenzialità del «significato esterno» dei loro testi, in particolare all'inadeguatezza di qualunque loro formalismo, non consente di portare molto avanti questo confronto.

Per Brouwer e la sua scuola intuizionista l'oggetto dell'intuizione originaria è la serie dei numeri interi 0, 1, 2, 3, ..., inoltre considerata non come realizzata, ma in divenire per mezzo della legge di passaggio da n ad $n+1$. Di conseguenza l'induzione è assunta come il modello fondamentale di tutte le costruzioni accettabili in matematica. Di per sé le costruzioni devono essere realizzate con metodi finitistici a partire da oggetti finiti e l'esistenza di un oggetto matematico è intesa come la possibilità di costruirlo. Ciò esclude le dimostrazioni di esistenza «per assurdo», non accompagnate da una rappresentazione dell'oggetto corrispondente, nei casi in cui occorra sceglierlo in un insieme potenzialmente infinito. È questo il noto principio di inaccettabilità della legge del terzo escluso nei domini infiniti.

Il continuo di Brouwer è difficilmente esplicabile nei termini classici di «immediato divenire libero», ad esempio delle frazioni decimali, non attualmente, ma solo potenzialmente infinite; nello stesso tempo, tuttavia, la scelta delle successive cifre decimali non è sottoposta ad alcuna legge.

Kronecker fu il primo predicatore delle limitazioni che questo punto di vista pone alla matematica, e uno dei piú influenti avversari di Cantor. Brouwer e i suoi successori dedicarono molti sforzi al chiarimento dei limiti della matematica e della logica intuizioniste, sviluppando i loro principî positivi. Tuttavia essi, in particolare Heyting, sottolineavano costantemente il primato

dell'interpretazione interna dei testi matematici e le limitate possibilità di trasmetterla in teorie assiomatiche intese in modo univoco. Con queste posizioni filosofiche, la mancanza di preparazione ad accogliere l'interpretazione interna della teoria degli insiemi nello spirito di Cantor può risultare infondata: molti matematici accettano nel suo complesso l'interpretazione interna della teoria degli insiemi. Per le motivazioni gli intuizionisti si appellano ad argomenti psicologici a difesa del primato dell'idea di suddivisibilità, discontinuità e, pare, numerabilità potenziale quale unico tipo legittimo di infinità. Agli intuizionisti non sono estranei gli appelli al significato «realistico» della matematica; in uno dei suoi articoli Brouwer afferma che l'applicazione del principio del terzo escluso nelle scienze naturali dipende dalle proprietà di finitezza e discretezza del mondo fisico.

Il costruttivismo della scuola di Markov cominciò a costituirsi alla fine degli anni '40, come variante dell'intuizionismo, mediante l'utilizzazione del concetto di oggetto costruibile e algoritmo, precisato in base alla teoria delle funzioni ricorsive o di qualche sua variante. Senza entrare qui nei particolari delle definizioni, verranno indicate soltanto alcune posizioni fondamentali di questa scuola.

La «realtà» dei costruttivisti è data dai testi su alfabeti finiti, sottoposti a processi di elaborazione che sono algoritmici, in particolare anch'essi descritti su un alfabeto finito.

L'infinità potenziale è ammessa come astrazione di testi, a priori di lunghezza non limitata, e di algoritmi che lavorano per un tempo a priori non limitato; l'infinità attuale viene rifiutata.

L'esistenza di un oggetto matematico è trattata come la sua esibizione, oppure come la possibilità di costruirlo algoritmicamente. La sottigliezza essenziale consiste nel fatto che, secondo Markov, la possibilità di costruzione algoritmica di per sé può essere stabilita con metodi classici e non ottenuta esclusivamente esibendo questa costruzione. In particolare in queste dimostrazioni si può utilizzare la riduzione all'assurdo.

Il continuo costruttivo è rappresentato da una coppia di algoritmi, il primo dei quali fornisce la successione numerabile di Cauchy dei numeri razionali, mentre il secondo fornisce chiare valutazioni sulla velocità del suo andamento (in senso classico). Perciò, dal punto di vista della matematica classica, è numerabile; oltre a ciò, *a*) la proprietà della coppia di algoritmi di definire un numero reale è algoritmicamente non decidibile, *b*) la proprietà di due di questi algoritmi di definire lo stesso numero non è decidibile; cosí pure non è decidibile la proprietà di rappresentare un dato numero. Per questi motivi l'analisi costruttiva si distingue nettamente da quella consueta. Ad esempio, al posto delle funzioni di variabile reale si considerano gli algoritmi che trasformano un numero reale costruttivo in un altro numero del genere. Ma allora una semplice funzione quale $\operatorname{sgn} x (= -1$ per $x<0$, $=0$ per $x=0$, $=+1$ per $x>0)$ non esiste nell'analisi costruttiva, perché altrimenti sarebbe risolubile la proprietà di un numero di «essere zero». In effetti tutte le funzioni costruttive sono continue nel senso classico della parola.

Quella «realtà» con la quale la matematica classica confronta le proprie costruzioni, per la maggioranza appare come una specie di «realtà» neoplatonica del mondo delle idee. Dal punto di vista del materialismo filosofico, si tratta della realtà delle coscienze individuali che riflettono il mondo esterno, sublimata ed estetizzata; dal punto di vista del platonismo, la realtà è indipendente dalle coscienze e si trova con queste in un accordo predisposto. È possibile anche considerare questo mondo come la parte comune alle proiezioni verso l'esterno dei significati interni della matematica nelle coscienze individuali. Una rappresentazione di unità di questo mondo può sussistere finché la voce discorde di eminenti matematici non dimostri il contrario.

È difficile tuttavia che qualsiasi argomento sulla semantica del concetto di insieme in matematica possa ritenersi definitivo. Una sua analisi attenta mostra un'estrema povertà del materiale scientifico riportato.

Ad esempio, criticando l'infinità attuale di Cantor delle essenze distinguibili, dalla posizione «realistica», si potrebbe osservare che essa non assolutizza tanto l'esperienza di uso del finito, quanto la fisica quotidiana in senso corrente, per la quale il mondo consiste di cose separate, che si possono contare, ordinare, riunire, ecc. La fisica quantistica suggerisce allora astrazioni del mondo di un tipo completamente diverso, nel quale un «insieme» di elettroni, ad esempio, consiste di essenze diverse ma non separabili; nel quale la logica non riflette la logica del discorso comune, orientato verso il macromondo, ma non coincide con la logica dell'intuizionismo, e in generale con nessuna delle versioni proposte dai filosofi della matematica. L'infinità del micromondo è «interna» ed «estesa», coi suoi aspetti probabilistici e relativi ai campi, assolutamente dissimile da quella di Cantor e definitivamente non riducibile a una successione di «atti elementari di suddivisione»; è la stessa cosa considerarla attuale piuttosto che potenziale. Il predicato stesso «appartenere a» ha nel micromondo uno status dubbio.

D'altro lato, anche le considerazioni psicologiche di Brouwer, Heyting e altri si richiamano soprattutto all'autoosservazione elementare e non prendono affatto in considerazione i risultati della psicologia scientifica e della neurofisiologia. Si può dire che un'attenta osservazione agli aspetti costruttivi della matematica ha permesso di isolare il concetto di macchina di Turing universale come «minimo denominatore» di tutte le coscienze individuali. Il suo fondamento fisiologico è l'attività discreta di movimento che utilizza i piú semplici recettori, ad esempio, tattili.

L'infinità potenzialmente numerabile degli intuizionisti e dei costruttivisti è un'infinità accessibile a questo denominatore. Tuttavia l'intuizione geometrica, sulla quale è fondata l'infinità del continuo, viene completamente ignorata in questo approccio. Il suo fondamento fisiologico è l'attività di movimento dello stesso piano ampliato, che si basa sui procedimenti complessi di elaborazione dell'informazione per mezzo di recettori visivi e di corrispondenti zone corticali. La descrizione neurofisiologica di questi procedimenti è appena all'inizio, ma è del tutto chiaro che concezioni semplificatrici dell'intuizione matematica, proposte a scopo polemico, sono lontane dalla realtà.

5.3. Aspetto esterno della semantica.

È complementare all'aspetto interno, nello stesso senso in cui la descrizione delle regole d'uso della parola 'insieme' è complementare alla sua definizione secondo Cantor. Concependo il testo matematico come un modello «esterno», il matematico può costruire le sue traduzioni in altri sottolinguaggi (dalle formule ai grafici e viceversa), costituire diverse abbreviazioni del testo o, al contrario, ricostruire passaggi saltati nelle dimostrazioni; sperimentare variazioni delle premesse; ecc. Il «significato esterno» di un testo matematico è il suo significato d'uso per il matematico di professione, quando scrive un articolo, insegna la matematica agli studenti o comunica coi colleghi. Il procedimento di interpretazione esterna non è obbligato, e raramente ciò accade, ad essere un'informazione completa, precisa, lineare sui concetti fondamentali (cosí è solo il suo modello nella teoria formale). Esso contiene tortuosità, strade senza sbocco e ostacoli. L'interpretazione della retta euclidea come modello dei numeri reali, e dei numeri reali come insieme dei punti della retta, non definisce nessuno di questi concetti, ma chiarisce il significato di entrambi.

L'interpretazione «realistica» della funzione delta di Dirac ($\delta(x)=0$ per $x\neq 0$, $\delta(0)=\infty$, $\int\delta(x)\,dx=1$), ad esempio, per un fisico, è quella di una massa puntuale con densità infinita. Per un matematico una variante piú astratta di questa rappresentazione si rivolgerà piuttosto alla categoria della semantica «interna». Soltanto con la possibilità di concepire la funzione delta in maniera «esterna», come funzionale sullo spazio delle funzioni a supporto compatto e racchiuderla in tal modo nei formalismi dell'analisi, in quanto termine ben definito, e nelle teorie matematiche generali in quanto autentico oggetto insiemistico, il matematico si sente completamente tranquillo. Tuttavia il fisico può ritenersi completamente soddisfatto dal formalismo esterno, in questo caso, senza postulare affatto la sua riducibilità alla teoria degli insiemi.

La componente filosofica delle concezioni formalistiche della matematica è connessa proprio con l'assolutizzazione del suo aspetto esterno, quando il contenuto viene completamente identificato con la forma. Con la cristallizzazione della matematica formale nei lavori di Hilbert e della sua scuola, si è di nuovo legati alle difficoltà della teoria degli insiemi.

Lo spirito innovatore di Cantor si è espresso prima di tutto nella formazione di una teoria intuitiva degli «ordini di infinità». Dal punto di vista linguistico ciò ha significato la confusa introduzione di alcuni assiomi di esistenza degli insiemi (fra cui l'assioma sull'insieme delle parti e l'assioma di scelta), l'utilizzazione di costruzioni linguistiche di nuovo tipo (come la definizione di ordinabilità totale) e, infine, la libera applicazione della logica del finito nel nuovo mondo infinitario (in particolare la legge del terzo escluso).

Questa valutazione del lavoro di Cantor, come si intende, si è resa possibile solo a posteriori; per esso, e per molto tempo dopo di esso, al primo posto si è situata la semantica dell'infinità attuale di tipo neoplatonico, anche se, probabilmente, non portata fino in fondo. In ogni caso, le affermazioni e i problemi

sull'esistenza dei vari insiemi furono intesi come aventi un significato comune a tutti. Gli stessi insiemi, di cui si affermava l'esistenza, potevano essere assegnati indicando le proprietà del loro elemento generico o costruendoli a partire da qualche insieme già dato.

Il trasferimento dell'accento dal «significato» al «testo» fu richiamato proprio dalla critica di questa semantica, sia dalla scoperta di paradossi, sia dall'inaccettabilità intuitiva per molti del teorema di Zermelo sulla possibilità di ordinare totalmente qualunque insieme. Il punto di vista di Hilbert consisteva nel fatto che non tutti i concetti e le considerazioni matematiche debbano essere immediatamente interpretabili in termini finitistici. Quelle che non ammettono una simile interpretazione sono costruzioni linguistiche «ideali», che sono aggiunte al sistema per semplificare le dimostrazioni dei teoremi, standardizzare il linguaggio, ecc. La loro ammissione nel sistema, tuttavia, richiede il rispetto di determinate regole, la principale delle quali è la non-contraddittorietà del sistema. La non-contraddittorietà deve essere garantita inizialmente per mezzo di un'accurata ricerca sulla sintassi del linguaggio, la quale deve essere condotta senza fare ricorso all'infinità attuale. Il suo risultato principale fu la varietà dei modelli formali di logica e di matematica sorti in conseguenza di questo programma. In particolare la costituzione dei concetti matematici precisi di numerabilità e di algoritmo e di tutta la teoria delle funzioni ricorsive è strettamente legata al programma di Hilbert. L'impossibilità di realizzarlo nella sua totalità, scoperta da Gödel, costituisce un altro aspetto del problema. È difficile decidere se questa impossibilità confermi l'accettazione della teoria intuitiva degli insiemi, come la si conosce oggi, oppure la sua negazione; con gli stessi fondamenti, questa domanda si può già porre per l'aritmetica.

In ogni caso, la visione della matematica come di un sistema formale è molto produttiva, quando se ne comprendono le limitazioni. Essa dà la possibilità di ottenere una notevole comprensione reciproca nell'esame di concezioni cosí diverse quali la logica intuizionista e la fisica teorica, fornendo gli «schemi svuotati» dalle strutture complesse. La non-contraddittorietà di un sistema formale, la cui utilità è verificata sperimentalmente, cessa di essere il compito primario. La scelta di testi concretamente generati nel sistema, costituenti una parte assolutamente insignificante di tutti i testi ammissibili, si esegue comunque secondo regole non formalizzate che sono piú importanti della non-contraddittorietà formale, descritta nei termini di tutti i testi sottoposti a generazione. Il ruolo dei testi scelti è quello fondamentale, sebbene poco studiato, di mediare fra un cervello e un altro, fra un cervello e il mondo esterno come pure fra un cervello e se stesso (il testo è una memoria esterna; il testo getta un ponte fra l'intuizione geometrica e quella aritmetica di ciascuna coscienza individuale, intuizioni che si basano su differenti meccanismi fisiologici). In ultima analisi, l'aspetto principale del testo è la sua possibilità di partecipare a questi atti di mediazione, e solo lo studio di questa possibilità può risolvere l'enigma della matematica. La vecchia metafora di Victor Hugo che paragona un libro e una cattedrale ha un significato profondo se la strutturabilità del testo costituisce la premessa piú importante del suo ruolo sociale.

Tutti questi problemi contribuiscono all'umanizzazione della matematica; insieme al movimento di matematizzazione della conoscenza umanistica, essi aiutano a ritrovare l'unità perduta delle due culture. Con una previsione sintetica, ci si può attendere una nuova concezione di «insieme» come uno dei piú eccezionali miti del nostro tempo. [JU. I. M.].

Artin, M.; Grothendieck, A.; e Verdier, J. L.
[1963-64] *Théorie des Topos et Cohomologie Etale des Schémas*, Séminaire de Géométrie Algébrique, vol. I. *Théorie des Topos*, Springer, Berlin - Heidelberg - New York 1972.

Bourbaki, N.
1954 *Eléments de mathématique*, I. *Théorie des ensembles* (fasc. 17), Hermann, Paris 1966³.

Burali-Forti, C.
1896 *Le classi finite*, in «Atti della R. Accademia delle Scienze di Torino», XXXII (1896-97), pp. 34-52.

Cantor, G.
1895-97 *Beiträge zur Begründung der transfiniten Mengenlehre*, in «Mathematische Annalen», XLVI (1895), pp. 481-512, e XLIX (1897), pp. 207-46; ora in *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*, Springer, Berlin 1932.

Fin dai suoi inizi la teoria degli insiemi si è posta come un elemento unificante a un livello di grande generalità per le diverse discipline **matematiche** (cfr. **calcolo, geometria e topologia, dualità**) e, attraverso una profonda crisi fondazionale, ha determinato il sorgere di diverse scuole e tendenze, relative ai problemi della verità in matematica, della dimostrabilità, dell'**infinito** e dell'**infinitesimale** (cfr. anche **vero/falso, verificabilità/falsificabilità**).

Oggigiorno il pensiero insiemistico è penetrato all'interno di tutta la **scienza** grazie alla creazione di formalismi efficienti (cfr. anche **fisica, forza/campo, energia**) e, posta attenzione alla definizione formale per non cadere in uno dei frequenti (e proficui) paradossi che ne hanno caratterizzato lo sviluppo, la nozione generale di insieme rappresenta il fondamento logico sul quale si eseguono le costruzioni e si conducono le dimostrazioni rigorose di maggior parte della matematica (cfr. **logica, assioma/postulato, deduzione/prova**).

Accanto alla concezione formale della teoria degli insiemi, dal punto di vista del contenuto, si scopre che in molti casi lo studio generale di una disciplina matematica si risolve nello studio di insiemi dotati di struttura e delle loro proprietà invarianti rispetto alle applicazioni compatibili con questa struttura (cfr. perciò **strutture matematiche, applicazioni, funzioni, invariante, variazione, continuo/discreto, simmetria**).

) finito. Questa li-
o (cfr. **qualità**/
la grandezza
un profilo
nalitiche
olice/
ulla

Razionale/algebrico/trascendente

1. *Introduzione.*

1.1. I termini del discorso.

I termini nel titolo di questo articolo si applicano in primo luogo ai numeri e alle funzioni.

Si dice numero razionale il rapporto fra due numeri interi, cioè un numero commensurabile con l'unità. Si dice funzione razionale delle variabili $x_1, \ldots, x_n$ il rapporto di due polinomi in queste variabili.

Si dice numero algebrico la radice di un polinomio i cui coefficienti sono numeri razionali. Si dice funzione algebrica delle variabili $x_1, \ldots, x_n$ una funzione legata a $x_1, \ldots, x_n$ da una relazione polinomiale.

Si dice numero trascendente ogni numero non algebrico. Si dice funzione trascendente ogni funzione non algebrica.

Tutti i numeri e tutte le funzioni razionali sono algebrici, ma il viceversa non è vero. Il numero $\sqrt{2}$, radice dell'equazione $x^2-2=0$ è algebrico ma non razionale. Piú in generale, se $a>0$ è un numero razionale, $n\geq 2$ è un intero e qualche numero primo p entra, nel numeratore o nel denominatore della rappresentazione irriducibile di a, ad una potenza non divisibile per n, allora $\sqrt[n]{a}$ è un numero reale algebrico, ma non razionale. Ciò deriva senza difficoltà dal teorema sulla decomposizione unica dei numeri razionali positivi come prodotto di potenze intere di numeri primi (si veda il § 1.5 dell'articolo «Divisibilità» in questa stessa *Enciclopedia*, articolo cui si farà continuo riferimento in seguito).

Sono esempi di numeri trascendenti il rapporto tra la circonferenza e il diametro $\pi=3{,}1415927\ldots$ e il numero di Nepero $\sum_{n=0}^{\infty}\frac{1}{n!}=e=2{,}7182818\ldots$ (cfr. § 4). Sono esempi di funzioni trascendenti: 2^x, tutte le funzioni trigonometriche, $\log x$, la funzione di Bessel, ecc.

L'etimologia del termine 'razionale' è legata alla rappresentazione delle operazioni di divisione dell'unità in parti uguali e alla loro successiva moltiplicazione per una quantità, che portano a tutti i numeri razionali positivi o, che è lo stesso, alla costruzione delle relazioni di divisibilità intera dell'unità. Tutti i numeri che non è possibile ottenere in tale modo, come $\sqrt{2}$, i numeri negativi (anche se interi), $i=\sqrt{-1}$, sono stati chiamati in tempi diversi «impossibili», «immaginari», «irrazionali», ecc. Dal punto di vista moderno in questi nomi si riflette il fatto che lo sviluppo storico del formalismo per lavorare coi numeri precedeva lo sviluppo della stessa rappresentazione del numero, vale a dire le interpretazioni di questo formalismo, e questi numeri denotarono a lungo un'esistenza fantomatica. Nella terminologia moderna la parola 'irrazionale' si è

riempita di tutti i numeri non razionali e 'immaginario' o 'immaginario puro' coi numeri $\mathbf{R}\sqrt{-1}$.

Tutti i numeri razionali costituiscono il campo $\mathbf{Q}$; tutti i numeri algebrici costituiscono il campo $\bar{\mathbf{Q}}$; i numeri trascendenti $\mathbf{C}\setminus\bar{\mathbf{Q}}$ non costituiscono un campo. Analoghe affermazioni valgono per le funzioni. Si ricordi che un campo è un insieme dotato delle operazioni di somma e di prodotto e degli elementi particolari 0 e 1; tutta questa struttura soddisfa gli assiomi dell'algebra elementare, validi per i numeri.

Nella matematica dei nostri giorni, la matematica dei razionali (degli algebrici, dei trascendenti) costituisce un capitolo piuttosto rilevante dell'algebra e della teoria dei numeri. In algebra questi termini si applicano, nel senso piú ampio della parola, in relazione a elementi di campi e di anelli, non necessariamente numerici o funzionali, e acquistano un senso relativo (si può parlare di elementi razionali, algebrici o trascendenti di qualche estensione di un anello, o di un campo, in relazione a un anello, o un campo, fondamentale). Nella teoria dei numeri si studiano in dettaglio l'aritmetica dei campi e degli anelli di numeri algebrici, i problemi di dimostrazione di trascendenza di varie classi di numeri, i problemi di approssimazione dei numeri trascendenti coi numeri algebrici, degli algebrici coi razionali, ecc. La parte principale di questo articolo (§§ 2-4) è dedicata all'esposizione di alcuni concetti e di alcuni risultati fondamentali di queste teorie.

La suddivisione dei numeri in razionali, algebrici e trascendenti riflette una classificazione dei numeri in base alla loro complessità. Questa classificazione ha una lunga storia, inseparabile dalla storia dello stesso concetto di numero e, in ultima analisi, dalla storia di tutta la matematica, in cui svolgono un ruolo fondamentale i concetti di numero e di figura geometrica. Si tracceranno qui brevemente le tappe di questa storia, organizzando l'esposizione intorno ad alcune idee centrali, di significato matematico generale. Fra queste idee rientrano: i linguaggi per rappresentare i numeri e le operazioni su questi; l'interpretazione semantica dei numeri in termini di misure e l'astrazione geometrico-numerica del continuo; i problemi del finito/infinito, del continuo/discreto e dell'informazione.

1.2. Il concetto di numero, linguaggio ed algebra.

Dai dati della linguistica e dell'etnografia si possono estrarre informazioni indirette sui primi periodi della storia delle rappresentazioni numeriche, relative ancora al periodo orale. In particolare il periodo «prematematico» è riflesso nei seguenti aspetti del sistema di denominazione dei numeri nelle diverse lingue naturali.

a) In alcune lingue primitive esistono denominazioni solo per numeri naturali piccoli, i rimanenti sono denotati con l'unica parola 'molto'. Cosí nelle lingue papuane ende, kati, kamoro, esistono i due numerali propri 1 e 2, fino a 20 si conta con le dita delle mani e dei piedi, a 20 la serie naturale «finisce». Ovviamente il sistema {1, ..., 20, molto} può essere formalizzato senza difficoltà,

con leggi di composizione naturali, come un «piccolo universo» non-contraddittorio della matematica. Inoltre questa, e analoghe, serie naturali sono riprese in considerazione da una scuola sui fondamenti della matematica (gli «ultra-intuizionisti») con uguale, se non maggiore, diritto rispetto alla serie naturale standard della matematica classica. Tuttavia non si può sopravvalutare il significato di questa raffinata ripresa di coscienza per i primi stadi e rifiutare la potente idea di prosecuzione infinita, potenziale o anche attuale, della serie dei numeri interi.

b) In alcune lingue si conservano varie serie di nomi dei naturali per contare oggetti di diversa natura (dritti, curvi, animati, inanimati, ecc.). Ciò testimonia del lungo periodo di formazione delle idee sui numeri come di strumenti adatti a contare «quanto si vuole». La coscienza umana per lungo tempo non è stata pronta a riunire in una «classe di equivalenza» insiemi arbitrari (anche se piccoli) di uguale potenza; quest'idea applicata agli insiemi infiniti è una conquista della matematica ottenuta con i lavori di Cantor. Traccia di simili tipi diversi per contare si è conservata nel cinese moderno, dove esiste un unico sistema di numerali, che è tuttavia completato da un evoluto sistema di particelle numerali utilizzate con sostantivi di varie classi, del tipo *kuài* 'pezzo', *ge* 'cosa', *běn* 'radice', ecc.

c) In numerose lingue si osservano diverse radici dalle quali si formano corrispondenti numerali d'ordine e di quantità (cfr. *uno/primo*, *duo/secundo* in latino). In queste testimonianze si può vedere il primo sorgere dell'idea di ordine (distinta da quella di quantità) che si costituirà come un concetto matematico autonomo molto piú tardi (si vedano i «cardinali» e gli «ordinali» di Cantor e le strutture d'ordine di Bourbaki).

I primi testi che ci sono giunti (babilonesi, egiziani) riflettono già un quadro evoluto delle conoscenze matematiche, in particolare la nascita di un linguaggio di notazioni matematiche abbastanza nettamente distinto dal linguaggio naturale. In esso il posto principale è occupato dal sistema per denotare i numeri e le operazioni su questi. La preistoria del sistema posizionale di notazione di tipo moderno è basata sull'idea di contare unità sempre piú grandi. Queste unità possono essere potenze di uno stesso numero (base del sistema posizionale) ma non è necessario. Cosí nel sistema cronologico dei Maya le unità sono 1, 18, 360 e poi $18\cdot 20^i$ (si vedano le tracce della numerazione in base 20 nei numerali francesi del tipo *quatre-vingt-six*). L'unità d'ordine superiore è notata con un simbolo speciale, che all'inizio può dipendere dal numero d'ordine (presso gli Egiziani, i Greci, i Romani). Quando questo simbolo cessa di dipendere dal numero d'ordine e il successivo è determinato soltanto dalla posizione nell'allineamento allora, per una notazione non ambigua, diventa necessario il simbolo zero. Si ritiene che la sua comparsa sia abbastanza lontana; al termine della tradizione babilonese l'assenza di unità di un dato ordine viene denotata con uno zero soltanto a metà della scrittura. L'idea che il simbolo zero non sia soltanto un segno, ma la notazione di un autentico numero, compare molto piú tardi ed è attribuita agli Indiani, dai quali l'avrebbe adottata la matematica europea attraverso l'influenza araba. La notazione posizionale contiene già il ger-

me della teoria delle operazioni algebriche sui numeri: la notazione richiede il prodotto di unità di ordine successivo per il numero di questa unità e la somma dei risultati.

Ulteriori sviluppi delle rappresentazioni numeriche sono legate all'uscita dal dominio dei numeri naturali.

Alla matematica babilonese è attribuito lo sviluppo dell'apparato di calcolo strettamente legato ai bisogni delle costruzioni e dell'economia, il quale ha condotto 1) all'abitudine ad operare con numeri non interi e relativamente grandi, 2) all'uso sistematico della prima operazione non razionale: l'estrazione di radice quadrata, e alla scoperta che questa è sufficiente per risolvere le equazioni quadratiche e biquadratiche. Nella maggior parte dei testi noti si estrae la radice quadrata solo dei quadrati perfetti; in casi eccezionali si usano approssimazioni abbastanza grossolane. Numeri negativi (compresa la loro interpretazione semantica come «debiti») si trovano nei documenti indiani. Tuttavia le regole formali per manipolarli sono di molto successive al sorgere del «modello standard» di continuo numerico, di modo che, in accordo con Bourbaki [1960], «in pieno XVIII secolo, d'Alembert, discutendo la questione nell'*Encyclopédie*... si perde improvvisamente d'animo dopo una colonna di spiegazioni piuttosto confuse, e si accontenta di concludere che "le regole delle operazioni algebriche sulle quantità negative sono generalmente ammesse da tutto il mondo ed accolte generalmente come esatte, quale che sia l'idea che si attribuisce a tali quantità"» (trad. it. p. 30).

Non va assunto un atteggiamento ironico verso queste difficoltà, le quali, come mostra la storia, sorgono inevitabilmente ad ogni livello del pensiero matematico (si veda l'articolo «Insieme» in questa stessa *Enciclopedia*); le si deve considerare tappe di una graduale presa di coscienza dell'idea fondamentale di separazione del linguaggio matematico dalla sua interpretazione (oggi essenzialmente insiemistica) e dell'idea di non-contraddittorietà del formalismo, come sostituti del concetto di «esistenza» degli oggetti matematici, formulata chiaramente nello storico manifesto di Hilbert del 1900.

In effetti, come è stato già ricordato, i numeri negativi e quelli complessi, fino al XVIII-XIX secolo figurano in matematica soprattutto come «interpretazioni potenziali» di qualche formalismo, piuttosto che come enti matematici definiti per se stessi. Questo formalismo consiste nel considerare espressioni con radici dei vari ordini, anche «ripetute», e nell'elaborazione di alcune regole per agire su esse. In questo modo occorre considerare, ad esempio, le formule di Bombelli (1572) $\sqrt[3]{2+\sqrt{-121}}=2+\sqrt{-1}$, di Leonardo da Pisa $\sqrt[3]{16}+\sqrt[3]{54}=\sqrt[3]{250}$, o, infine, le formule generali di Del Ferro (1545) per le radici dell'equazione $x^3+ax=b$:

$$x=\sqrt[3]{\frac{b}{2}+\sqrt{\left(\frac{b}{2}\right)^2+\left(\frac{a}{3}\right)^3}}+\sqrt[3]{\frac{b}{2}-\sqrt{\left(\frac{b}{2}\right)^2+\left(\frac{a}{3}\right)^3}}.$$

Come è noto, nel caso $(b/2)^2+(a/3)^3<0$ occorre calcolare la radice reale, per mezzo di queste formule, passando attraverso grandezze complesse, le quali si

presentano come oggetti di calcolo «ideali», vale a dire «non interpretabili».

Ovviamente quest'idea fu compresa completamente molto piú tardi e, fino alla formazione della teoria degli insiemi da un lato e della metamatematica dall'altro, non poté essere totalmente assimilata. Tuttavia la compatibilità interna dei buoni formalismi al di fuori del dominio della loro applicazione originaria è un fenomeno continuamente ricorrente nella storia della matematica e uno dei maggiori fattori di progresso. Uno degli esempi piú moderni è l'integrale di Feynmann (si veda il § 3.3 dell'articolo «Dipendenza/indipendenza» in questa stessa *Enciclopedia*).

Occorre anche tener conto del fatto che ogni formalismo è «immediatamente accessibile» e che è un concetto finito, mentre le interpretazioni sono sottoposte a variazioni storiche e soprattutto sono semanticamente multiformi.

1.3. Il concetto di numero, geometria e teoria degli insiemi.

Una seconda corrente di idee risale alla matematica greca, nella quale sorge il concetto di continuo numerico come astrazione delle figure geometriche e della teoria delle relazioni. Bourbaki sottolinea che sia la parola sia l'idea di numero, nella teoria di Eudosso esposta da Euclide, si rivolgono esclusivamente ai numeri interi maggiori di 1. I numeri sono considerati come un sistema con somma e prodotto. I rapporti fra numeri, cioè i numeri razionali (positivi) nella terminologia moderna, sono introdotti come operatori che agiscono su domini di grandezze di una stessa natura (lunghezze, aree, ecc.): il numero ab^{-1} è l'operatore che consiste nel costruire a parti b-esime di una data grandezza. Questi operatori possono moltiplicarsi, ma non comporsi. Grandezze della stessa natura possono essere confrontate e soddisfano il noto «assioma di Archimede»: la somma di un numero abbastanza alto di esemplari di una data grandezza può diventare grande quanto si vuole. La struttura algebrica dei numeri reali, in tal modo, rimane suddivisa in un dominio degli operatori (prodotto) e in un dominio delle grandezze stesse (somma). Il concetto di ordine svolge un ruolo fondamentale per dare un senso alla relazione fra grandezze incommensurabili, come la diagonale di un quadrato e il suo lato, e ciò permette di considerare Eudosso come un precursore della teoria dei numeri reali secondo Dedekind. Il concetto di continuo topologico si costituisce anch'esso in ricerche come il problema della duplicazione del cubo. La costruzione di $\sqrt[3]{2}$, necessaria alla sua soluzione, si riduce alla ricerca del punto d'intersezione di due curve, le quali si costruiscono con un processo meccanico idealizzato. L'esistenza di un punto d'intersezione come conseguenza del fatto che la seconda curva contiene punti che si trovano da parti opposte della prima, evidentemente, era considerata come ovvia. È noto il ruolo che svolge il corrispondente teorema nel sistema della matematica contemporanea.

Infine la rappresentazione del continuo geometrico come ricettacolo di relazioni numeriche fu un'idea precorritrice della geometria analitica di Descartes, da un lato, e della teoria degli insiemi dall'altro. Dal punto di vista psicologico, la percezione visiva che sta alla base delle rappresentazioni geometriche si con-

figura come l'unico canale di passaggio che collega l'uomo col mondo esterno. A queste sue particolarità è legata la possibilità di astrazione del continuo in quanto oggetto in cui un insieme di posizioni di un punto, fra due punti dati, possiede le proprietà di continuità e di estensione illimitata in profondità – un'idea che solo indirettamente e a fatica si formula nel linguaggio delle successioni potenzialmente infinite di simboli numerici, sviluppati rispetto a un dato sistema di numerazione. Soprattutto, l'infinità del continuo si concepisce piú facilmente come assegnata attualmente.

L'eliminazione del dualismo numeri discreti / relazioni continue fra grandezze nel significato generale dei numeri reali, richiese ancora molti secoli di sviluppo. Tuttavia l'accessibilità immediata del continuo alla intuizione, seppure limitata, svolse un ruolo importante nell'assegnazione di uno status riconosciuto ai numeri complessi, dopo la scoperta della loro interpretazione come punti del piano complesso. Per questo, teorie come quella dei limiti di Cauchy, la costruzione dei numeri reali secondo Dedekind e Cantor, l'algebra complessa e l'analisi, possono essere considerate il punto d'incontro nella formazione di un formalismo che descrive l'idealizzazione di un dato quasi «fisico», «a priori», della realtà. L'unione di questi due indirizzi si verificò sulla base della teoria degli insiemi.

Per tornare al tema immediato di questo articolo, il termine 'trascendente' appartiene a Leibniz. Fu introdotto, evidentemente, in relazione alle funzioni. In particolare Leibniz dimostrò (1682) che $\sin x$ non è una funzione algebrica di x e risolvette allo stesso tempo il problema di Gregori (1667): l'area di un settore circolare non si esprime algebricamente per mezzo del raggio e della corda.

L'esistenza di numeri trascendenti fu dimostrata per la prima volta da Liouville (1844) per mezzo della seguente, considerevole, osservazione, che dette inizio alla moderna teoria dell'approssimazione. Sia α una radice irrazionale del polinomio irriducibile $f(x)$ di grado n a coefficienti interi. Allora, per ogni numero razionale p/q si ha

$$f\left(\frac{p}{q}\right)\neq 0 \Rightarrow \left|f\left(\frac{p}{q}\right)\right| \geq \frac{1}{q^n}.$$

D'altro lato se p/q è abbastanza vicino a α, allora $f(p/q)\approx[(p/q)-\alpha]f'(\alpha)$ e $f'(\alpha)\neq 0$. Da qui si conclude facilmente che $|\alpha-(p/q)|\geq c/q^n$ per una costante opportuna $c>0$ e per ogni numeri razionale p/q. Di conseguenza, se per il numero dato $\alpha\in\mathbf{R}$ e per ogni $c>0$, $\nu>0$ esistono $p,q\in\mathbf{Z}$ tali che $|\alpha-(p/q)|<(c/q^\nu)$, allora α non può essere algebrico. Un esempio semplice $\alpha=\sum_{n=0}^{\infty}\frac{1}{2^{2^n}}$: con le somme parziali $\sum_{n=1}^{N}\frac{1}{2^{2^n}}$ questo numero viene approssimato molto bene.

La trascendenza di numeri come e e π, tuttavia, non si può dimostrare con il metodo di Liouville. La trascendenza di e fu stabilita da Hermite (1873) e per π da Lindemann (1882), dopo che aveva dimostrato in tal modo l'impossibilità di quadrare il cerchio per mezzo di riga e compasso – un antico problema che

serviva da stimolo per queste ricerche. Un accenno a questa dimostrazione è riportato piú avanti, nel § 4. Nella nostra epoca la teoria dei trascendenti ha molti risultati e non meno problemi irrisolti. Ad esempio la questione della «costante di Eulero» $C=\lim_{N\to\infty}\left(\sum_{n=1}^{N}\frac{1}{n}-\log N\right)$ rimane irrisolta. La maggior parte dei problemi relativi alla trascendenza di numeri concreti è riferita a numeri assegnati «analiticamente», vale a dire come valori di varie funzioni, come loro zeri, ecc. Di regola questi numeri possono essere calcolati algoritmicamente con la precisione che si vuole. Allo stesso tempo essi costituiscono solo una piccola parte di tutto il continuo, come è stato dimostrato in una famosa dimostrazione di esistenza dei numeri trascendenti da parte di Cantor.

La dimostrazione di Cantor è basata sulla sua teoria delle potenze degli insiemi infiniti e consiste nella constatazione del fatto che l'insieme dei numeri algebrici è numerabile, mentre l'insieme di tutti i numeri reali (e complessi) ha la potenza del continuo, nettamente maggiore. La considerazione di Cantor presenta il difetto di non essere effettiva, vale a dire stabilisce l'esistenza (e la potenza) dei numeri trascendenti senza la possibilità di costruirne «esplicitamente», cioè con i passaggi al limite consueti in matematica, anche uno solo di essi. Allo stesso tempo possiede il notevole pregio di generalizzarsi a qualunque classe numerica generata da un processo finito, o almeno descrivibile in un linguaggio formale di tipo finito. Fra le classi piú importanti di questo tipo rientra il «continuo costruttivo».

1.4. Il concetto di numero e l'informazione.

Il continuo costruttivo (in diverse varianti) fu proposto dai moderni critici dei fondamenti della matematica come un'alternativa al continuo classico all'interno della «matematica alternativa», che non utilizza il concetto di infinità effettiva. Elementi del continuo costruttivo di Markov sono classi di successioni di Cauchy effettivamente computabili di numeri razionali (a_n) per le quali esistono le funzioni effettivamente computabili $f:\mathbf{Z}\to\mathbf{Q}$, $f(n)\to 0$ per $n\to\infty$, e $g:\mathbf{Z}\to\mathbf{Z}$ tali che $|a_n-a_m|<f(M)$ quando $n,m>g(M)$.

Si riuniscono nella stessa classe le successioni che convergono allo stesso numero, in senso classico.

La complessità dei problemi strutturali del continuo costruttivo è legata già al fatto che il problema di identificare due numeri reali, assegnati mediante algoritmi di calcolo delle loro successioni di Cauchy (a_n) e (b_n), non è algoritmicamente risolubile. Ma anche rinunziando alla proibizione di definizioni non effettive, il continuo costruttivo, in quanto oggetto della matematica classica, è poco studiato. Fra l'altro si tratta di un oggetto assolutamente fondamentale, in quanto gran parte dei problemi di teoria dei numeri, sia algebrici sia trascendenti, si pone essenzialmente al suo interno. I numeri costruttivi sono numeri che portano una «quantità finita d'informazione» e come tali svolgono un ruolo particolare. I numeri non costruttivi si potrebbero chiamare «ultra-trascendenti»: ciascuno di questi numeri, in qualche senso del termine, non è dato da una

computazione finita. Naturalmente sono possibili generalizzazioni e varianti. Ad esempio, piú ampia del continuo costruttivo è la classe dei numeri α per i quali è derivabile nella teoria degli insiemi la formula $\exists!\alpha(P(\alpha))$, dove $P(x)$ è una formula del linguaggio della teoria degli insiemi, con la variabile libera x, che assume valore in **R**. In qualche senso si può considerare che questi numeri siano ancora descritti in maniera finita, anche se non necessariamente computati in maniera finita.

L'argomento di Cantor mostra che tutti questi «continui» sono numerabili. Il loro studio è un compito del futuro.

2. *Anelli, campi e teoria di Galois.*

2.1. La costruzione dei razionali.

Il campo dei numeri razionali **Q** si ottiene dall'anello dei numeri interi **Z** con l'aggiunta di tutte le frazioni pq^{-1}, vale a dire dei risultati della divisione formale di numeri interi, non realizzabile in **Z**. Il campo dei numeri algebrici $\bar{\mathbf{Q}}$ si ottiene dal campo **Q** con l'aggiunta delle radici dei polinomi a coefficienti razionali, che di per sé non appartengono a **Q**. Anche il campo dei numeri reali **R** può essere descritto come il risultato che si ottiene da una «operazione infinita» – il passaggio al limite per le successioni formate da elementi di **Q**. L'assiomatizzazione di queste idee nell'algebra moderna conduce alla teoria degli anelli e dei campi astratti, alle loro estensioni, ai completamenti topologici, ecc. Il punto di partenza per lo studio dell'aritmetica degli anelli e dei campi numerici e funzionali, è costituito proprio dai concetti generali cui tale assiomatizzazione ha condotto.

2.2. Anelli commutativi e domini d'integrità.

Si ricordi che un anello è un insieme A con due operazioni di somma e di prodotto, con gli elementi distinti 0 e 1, i quali soddisfano agli abituali assiomi dell'algebra elementare. Ci si limita qui al caso di anelli con prodotto commutativo. Un anello si dice dominio d'integrità se non ha divisori dello zero, vale a dire se da $ab=0$ segue $a=0$ oppure $b=0$. Un anello si dice campo quando $A^*=A-\{0\}$ è un gruppo rispetto al prodotto. Ogni campo è un dominio d'integrità. In ogni anello commutativo A esiste un sottoanello minimo contenente l'unità: è costituito da tutti i multipli interi dell'unità ed è isomorfo a **Z** oppure a $\mathbf{Z}/m\mathbf{Z}$, $m\in\mathbf{Z}$, $m>0$. Se un dominio d'integrità contiene un sottoanello isomorfo a $\mathbf{Z}/m\mathbf{Z}$, allora m deve essere necessariamente un numero primo (altrimenti in $\mathbf{Z}/m\mathbf{Z}$ esisterebbe un divisore dell'unità), che si chiama caratteristica di A. Se $\mathbf{Z}\subset A$ si dice che la caratteristica è zero. Gli anelli **Z** e $\mathbf{Z}/p\mathbf{Z}$, con p primo, si dicono primi; ogni dominio d'integrità contiene un sottoanello primo. Se A e B sono due anelli e $A\subset B$, allora B si dice estensione di A.

2.3. Costruzioni di anelli.

La principale costruzione è l'aggiunta di una variabile indipendente. Per descriverla si ponga

$$B=\{(a_0, a_1, a_2, \ldots) \mid a_i\in A,\ \exists i_0,\ \forall i\geq i_0,\ a_i=0\}.$$

In altri termini B è l'insieme delle successioni infinite di elementi di A che, tranne un numero finito, sono nulli. S'introduca in B una struttura di anello commutativo ponendo $(a_i)+(b_i)=(a_i+b_i)$, $(a_i)\cdot(b_i)=\left(\sum_{k=0}^{j} a_{j-k}b_k\right)$. È facile verificare gli assiomi per B; in particolare $(0, 0, \ldots)$ è lo zero, mentre $(1, 0, 0, \ldots)$ è l'unità in B. Si ponga quindi $T=(0, 1, 0, \ldots)\in B$; allora $(a_i)=\sum_{i=0}^{\infty} a_iT^i$ (quasi tutti i «monomi» a_iT^i sono uguali a zero) e la scrittura di ogni elemento di B in questa forma è unica. Pertanto si può identificare l'anello B con l'anello $A[T]$ dei polinomi nella «variabile indipendente» T, a coefficienti in A. Iterando questa costruzione un numero finito di volte si ottiene l'anello dei polinomi in n variabili $A[T_1, \ldots, T_n]$. L'ordine con cui si aggiungono le variabili è indifferente, nel senso che aggiungendo in ordine diverso le stesse variabili gli anelli ottenuti sono isomorfi in maniera canonica. Ciò permette di definire l'anello dei polinomi $A[S]$ in un insieme qualsiasi di variabili S come $\bigcup A[S_i]$, dove $\ldots\subset S_i\subset S_{i+1}\subset\ldots$ è una successione d'insiemi finiti che esaurisce S (se S non è numerabile, bisogna usare gli ordinali e non solo i numeri interi come indici della successione). Si ha un'immersione insiemistica canonica di S in $A[S]$.

Se A è un dominio d'integrità, allora anche $A[S]$ è un dominio d'integrità. L'anello $A[S]$ è caratterizzato dalla seguente proprietà universale: per ogni omomorfismo di anelli $f: A\to B$ e ogni applicazione insiemistica $g: S\to B$, esiste un unico omomorfismo $h: A[S]\to B$, coincidente con f su A e con g su S.

Da qui segue che ogni dominio d'integrità A è isomorfo al quoziente di un anello di polinomi $A_0[S]$, dove $A_0=\mathbf{Z}$ oppure $\mathbf{Z}/p\mathbf{Z}$ è un anello primo con la stessa caratteristica di A. Affinché il quoziente $A/\mathfrak{i}$ dell'anello A rispetto all'ideale $\mathfrak{i}$ sia un dominio d'integrità, è necessario e sufficiente che l'ideale $\mathfrak{i}$ sia primo (cfr. il § 4.5 dell'articolo «Divisibilità» in questa stessa *Enciclopedia*).

2.4. Campo dei quozienti di un dominio d'integrità.

Sia A un dominio d'integrità e sia $S=(A\setminus\{0\})\times A$. È comodo denotare con $T(q,p)$ l'elemento di S corrispondente alla coppia (q,p). Si consideri il quoziente $K=A[S]/\mathfrak{i}$ rispetto all'ideale $\mathfrak{i}$ generato da tutti i polinomi di tipo $qT(q,p)-p$. Non è difficile vedere che si tratta di un campo e che l'applicazione canonica $A\to A[S]\to A[S]/\mathfrak{i}$ è un monomorfismo. Identificando A con la sua immagine in K, si può mostrare che ogni elemento di K si rappresenta nella forma ab^{-1}, $a,b\in A$, $b\neq 0$; in particolare l'immagine di $T(q,p)$ in K è pq^{-1}. Una costruzione equivalente di K è la seguente: si consideri lo stesso insieme S e

s'introduca in esso la relazione di equivalenza: $(q,p) \sim (q',p')$ se e solo se esistono degli elementi $r, r' \in A \setminus \{0\}$ tali che $(rq, rp) = (r'q', r'p')$. Si denoti con K il quoziente di S rispetto a questa relazione di equivalenza e s'introducano in esso le operazioni di somma e di prodotto: $\text{classe}(q,p) \times \text{classe}(q',p') = \text{classe}(qq', pp')$; $\text{classe}(q,p) + \text{classe}(q',p') = \text{classe}(qq', pq' + q'p)$. Si può dimostrare che questa definizione è corretta e K costituisce un campo contenente $A = \{\text{classi}(1,p)\}$ in cui $\text{classe}(q,p) = pq^{-1}$.

Il campo K costruito sopra si dice campo dei quozienti del dominio d'integrità A. L'immersione $A \subset K$ dell'anello A nel suo campo dei quozienti possiede la seguente proprietà universale: per ogni monomorfismo $f: A \to L$ dell'anello A nel campo L esiste un unico monomorfismo $K \to L$ coincidente con f su A.

Il campo dei quozienti dell'anello dei numeri interi $\mathbf{Z}$ coincide col campo dei numeri razionali $\mathbf{Q}$: si dice campo primo di caratteristica zero. Gli anelli primi $\mathbf{Z}/p\mathbf{Z}$ stessi sono campi. Si tratta di campi primi di caratteristica finita. Ogni campo possiede un sottocampo primo della stessa caratteristica.

Se K è un campo, $T_1, \ldots, T_n$ sono variabili, il campo dei quozienti dell'anello $K[T_1, \ldots, T_n]$ si denota con $K(T_1, \ldots, T_n)$ e si chiama campo delle funzioni razionali in $T_1, \ldots, T_n$ a coefficienti in K.

Entrambe le costruzioni del campo dei quozienti ammettono varianti e generalizzazioni nel caso in cui non si supponga che A sia un dominio d'integrità. La piú generale di queste è la costruzione dell'anello dei quozienti A_R rispetto a un sottomonoide (relativo al prodotto) $R \subset A$: l'anello A_R è costituito dai rapporti formali ar^{-1}, dove $a \in A$, $r \in R$, con le abituali regole per operare sulle frazioni. Si ha sempre un omomorfismo canonico $A \to A_R$: $a \mapsto a \cdot 1^{-1}$, ma questo può non essere un'immersione, se R contiene divisori dello zero.

Le operazioni di somma e di prodotto, e le loro inverse – sottrazione e divisione – sono tradizionalmente dette operazioni razionali. La teoria degli anelli e dei campi è la teoria generale delle «operazioni razionali» negli insiemi.

2.5. Campi ed estensioni di campi.

Siano K ed L due campi e sia $K \subset L$; in questo caso il campo L si dice estensione del campo K e si denota a volte con L/K. Un elemento $a \in L$ si dice algebrico su K se è radice di un polinomio a coefficienti in K. In caso contrario l'elemento $a \in L$ si dice trascendente su K. Un'estensione L del campo K si dice algebrica su K se tutti i suoi elementi sono algebrici su K. Un'estensione L si dice trascendente pura su K se è isomorfa al campo dei quozienti di un anello di polinomi $K[T_1, T_2, \ldots]$. In altri termini, L si ottiene aggiungendo a K degli elementi algebricamente indipendenti.

2.6. Grado di trascendenza.

Per ogni estensione di campi $K \subset L$ è possibile trovare un campo intermedio M, $K \subset M \subset L$, tale che l'estensione $K \subset M$ sia puramente trascendente e l'estensione $M \subset L$ sia algebrica. Il campo M non è univocamente determinato, tuttavia il numero di elementi trascendenti che occorre aggiungere a K per ottenere M dipende solo da L. Questo numero è detto grado di trascendenza dell'estensione $K \subset L$.

2.7. Grado delle estensioni algebriche.

Se l'estensione $K \subset L$ è algebrica, vale a dire il suo grado di trascendenza è nullo, allora la dimensione di L come spazio lineare su K si dice grado di L. Se questa dimensione è finita, l'estensione $K \subset L$ è necessariamente algebrica: ogni elemento $a \in L$ si può rappresentare con una matrice di prodotto per a rispetto a una opportuna base di L su K, e questa matrice, e con essa a, è la radice della sua equazione caratteristica. Tuttavia possono sussistere estensioni algebriche di grado infinito: sono tali ad esempio le chiusure algebriche dei campi primi.

2.8. Estensioni algebriche separabili e non separabili.

Sia $K \subset L$ un'estensione di campi e $a \in L$ sia algebrico su K. Allora a è radice di un polinomio f irriducibile su K e questo polinomio è univocamente determinato a meno di un fattore non nullo di K. L'elemento a si dice separabile su K se f non ha radici multiple. Un'estensione algebrica L/K si dice separabile se tutti gli elementi di L sono separabili su K. Quando la caratteristica del campo K è zero, tutte le sue estensioni sono separabili. Sono separabili anche tutte le estensioni algebriche dei campi primi e, piú in generale, dei campi perfetti di caratteristica p, vale a dire i campi K con la proprietà $K^p =_{\text{def}} \{a^p \mid a \in K\} = K$. Se $K^p \neq K$ l'estensione $K^p \subset K$ è algebrica e non separabile. Il motivo di ciò risiede nel fatto che se $t \in K$ e $t \notin K^p$, il polinomio $x^p - t^p$ è irriducibile su K^p ma tutte le sue radici coincidono, giacché $x^p - t^p = (x-t)^p$.

2.9. Estensioni finite di Galois.

Sia L un campo, G un gruppo finito di suoi automorfismi, cioè di applicazioni biiettive $L \to L$ che portano la somma di elementi nella somma delle loro immagini e il prodotto nel prodotto delle immagini. Si denoti con $L^G = K$ l'insieme dei punti uniti, vale a dire $K = \{x \in L \mid gx = x \text{ per ogni } g \in G\}$. Non è difficile vedere che K è un campo.

Le affermazioni che seguono costituiscono la parte centrale della teoria di Galois.

a) L'estensione $K \subset L$ è un'estensione finita, separabile, di grado $\text{Card}\, G$ (cioè il numero di elementi di G). Il gruppo G è il gruppo di tutti gli automorfismi di L che lasciano invariati tutti gli elementi di K.

b) Si faccia corrispondere ad ogni sottogruppo $H \subset G$ il campo dei suoi elementi uniti L^H. Allora la corrispondenza $H \mapsto L^H$ determina una biiezione fra l'insieme dei sottogruppi di G e l'insieme dei campi M, intermedi fra K e L. Questa biiezione inverte l'ordine rispetto all'inclusione: $H_1 \subset H_2 \Leftrightarrow L^{H_1} \supset L^{H_2}$.

Le estensioni $K \subset L$ del tipo descritto si dicono estensioni di Galois con gruppo di Galois G.

c) Nella situazione descritta l'estensione $K \subset L^H$ è un'estensione di Galois se e solo se $H \subset G$ è un sottogruppo normale e, se questa condizione è soddisfatta, il gruppo di Galois di $K \subset L^H$ è canonicamente isomorfo a G/H.

2.10. Esempi.

a) $K = \mathbf{Q}$, $L = \mathbf{Q}(\xi)$ dove ξ è una radice primitiva di ordine m dell'unità. Per ogni numero intero n, primo con m, l'applicazione $\xi \mapsto \xi^n$ induce un automorfismo di L su $\mathbf{Q}$, dipendente solo dalla classe dei resti di $n \bmod m$. Tutti gli automorfismi di questo tipo costituiscono un gruppo canonicamente isomorfo al gruppo delle classi di resti $(\mathbf{Z}/m\mathbf{Z})^*$ rispetto al prodotto. Il suo ordine è uguale a $\varphi(m)$ (funzione di Eulero) e tale è anche il grado $\dim_{\mathbf{Q}} L = [L : \mathbf{Q}]$. Pertanto $L/\mathbf{Q}$ è un'estensione di Galois.

b) Sia ancora $L = \mathbf{Q}(\xi)$ e $a \in \mathbf{Q}(\xi)$ sia un elemento tale che l'estensione $M = L(\sqrt[m]{a})$ abbia grado m su L, vale a dire il polinomio $x^m - a$ sia irriducibile su L. Allora per ogni numero intero n, l'applicazione $\sqrt[m]{a} \mapsto \xi^n \sqrt[m]{a}$ induce un automorfismo di M rispetto a L. Tutto il gruppo di questi automorfismi è isomorfo al gruppo μ_m delle radici di grado m dell'unità, e $L \subset M$ è un'estensione di Galois con gruppo μ_m.

c) Sia p un numero primo, $F_p = \mathbf{Z}/p\mathbf{Z}$, $q = p^m\ m \geq 1$; F_q è un campo finito con q elementi (cfr. ancora «Divisibilità», § 4.3). L'applicazione $\Phi : F_q \to F_q$, $\Phi(x) = x^p$ è un automorfismo del campo F_q; F_p coincide col campo dei suoi elementi uniti. Tutte le potenze di Φ costituiscono un gruppo ciclico di ordine m, il quale è il gruppo di Galois dell'estensione $F_p \subset F_q$.

d) Sia K un campo di caratteristica diversa da 2 e da 3, $f(X) = X^3 + aX + b$, $a, b \in K$. Si indichi con L il campo ottenuto aggiungendo a K tutte le radici di f. L'estensione $K \subset L$ è un'estensione di Galois con gruppo simmetrico S_3 di ordine 6 se e solo se il discriminante $\Delta = -4b^3 - 27b^2$ non è un quadrato nel campo K.

2.11. Caratterizzazione delle estensioni di Galois.

Un'estensione algebrica $K \subset L$ si dice normale se possiede la seguente proprietà: ogni polinomio di $K[T]$ irriducibile in K che ha una radice in L si decompone in fattori lineari in $L[T]$.

Un'estensione algebrica finita $K \subset L$ è un'estensione di Galois se e soltanto se è normale e separabile.

2.12. Estensioni generali di Galois.

Un'estensione algebrica $K \subset L$ si dice estensione di Galois se si può rappresentare come unione di estensioni finite di Galois. I risultati del § 2.8 si trasferiscono a questo caso se nel gruppo totale degli automorfismi di L su K

s'introduce la topologia di Krull, nella quale una base di intorni dell'unità è data dai sottogruppi di automorfismi che lasciano fisse le sottoestensioni finite L su K. La corrispondenza $H \mapsto L^H$ determina allora una biiezione fra i sottogruppi chiusi del gruppo di Galois dell'estensione $K \subset L$ e i campi intermedi fra K e L, non necessariamente di grado finito su K.

Se il campo K è perfetto e $\bar{K}$ è la sua chiusura algebrica, allora l'estensione $\bar{K}/K$ è un'estensione di Galois. Il suo gruppo di Galois $\mathrm{Gal}(\bar{K}/K)$ è il principale invariante del campo. Esso è computabile solo in alcuni casi. In particolare per $K = \mathbf{Q}$ (o un'estensione finita di $\mathbf{Q}$) la sua struttura è ben lontana dall'essere chiara. Numerosi risultati sul suo quoziente rispetto ai commutatori, cioè sul gruppo di Galois dell'estensione abeliana massimale del campo numerico K, sono oggetto della profonda teoria dei campi di classi, di cui si tratterà piú dettagliatamente nel prossimo paragrafo. Ecco due esempi in cui si può calcolare il gruppo $\mathrm{Gal}(\bar{K}/K)$.

a) $K = F_p$. Qui $\bar{K}$ è l'unione di tutti i campi F_{p^m}, $m > 1$. Il gruppo $\mathrm{Gal}(\bar{F}_p/F_p)$ è generato topologicamente dall'automorfismo di Frobenius $\Phi : x \mapsto x^p$. Ciò significa che è il completamento rispetto alla topologia di Krull del gruppo ciclico libero $\mathbf{Z} \simeq \{\Phi^m \mid m \in \mathbf{Z}\}$. Una base di intorni dello zero in questa topologia è costituita da tutti i sottogruppi del tipo $n\mathbf{Z}$, $n \in \mathbf{Z}$. Il completamento $\hat{\mathbf{Z}}$ rispetto ad essa è isomorfo a $\prod_p \mathbf{Z}_p$ – prodotto di gruppi d'interi p-adici rispetto a tutti i numeri primi p, con la topologia del prodotto delle topologie p-adiche.

b) $K = \mathbf{C}(t)$ sia il campo delle funzioni razionali in una incognita, a coefficienti complessi. Qui la descrizione del gruppo $\mathrm{Gal}(\bar{K}/K)$ si basa sulle seguenti considerazioni topologico-analitiche. Si rappresenti il campo K come campo delle funzioni meromorfe sulla sfera di Riemann X. Ogni estensione algebrica finita $L \subset K$ si può rappresentare, in maniera analoga, come campo delle funzioni meromorfe sulla superficie compatta di Riemann Y, rivestimento d'ordine finito di X. Questo rivestimento si ramifica in un numero finito di punti $P_1, \ldots, P_s \in X$, dipendenti da L. Una classificazione puramente topologica di questi rivestimenti li associa ai sottogruppi di indice finito del gruppo fondamentale $\pi_1(X - \{P_1, \ldots, P_s\})$ il quale è il gruppo libero (non commutativo) con $s - 1$ generatori. Ogni rivestimento topologico di questo tipo ha una struttura complessa naturale, di modo che ad esso si associa il corrispondente campo delle funzioni meromorfe L (cfr. l'articolo «Locale/globale» in questa stessa *Enciclopedia*, § 3). Come risultato il gruppo $\mathrm{Gal}(\overline{\mathbf{C}(t)}/\mathbf{C}(t))$ si può identificare col limite proiettivo $\varprojlim \pi_1(X - S)/H$, dove S varia lungo i sottoinsiemi finiti di X e H lungo i divisori normali di indice finito. $\mathrm{Gal}(\overline{\mathbf{C}(t)}/\mathbf{C}(t))$, in termini tecnici, è un «gruppo libero profinito con un insieme continuo di generatori».

Queste considerazioni geometriche si trasferiscono al caso di campi piú generali K, contenenti elementi trascendenti su sottocampi primi, almeno quando K ha un grado di trascendenza finito. Essi mostrano il profondo legame della teoria di Galois di questi campi con la topologia delle varietà. Inoltre la teoria di Galois del campo $F_p(t)$ e delle sue estensioni finite fornisce un buon esempio di «poligono di prova» per l'aritmetica dei numeri algebrici. Infatti,

da un lato la torre $F_p[t] \subset F_p(t) \subset \overline{F_p(t)}$ possiede proprietà algebriche prossime alle proprietà della torre $\mathbf{Z} \subset \mathbf{Q} \subset \bar{\mathbf{Q}}$. D'altro lato ad essa è applicabile una notevole tecnica algebro-geometrica nella quale si hanno analoghi per quasi tutti i metodi della topologia classica e della teoria delle funzioni su una superficie di Riemann, compreso la teoria della (co)omologia [si vedano i particolari in Šafarevič 1962].

3. *Aritmetica dei numeri algebrici.*

3.1. Numeri algebrici interi.

Si consideri inizialmente la seguente situazione generale: sia A un anello commutativo, K il suo campo dei quozienti, $K \subset L$ un'estensione algebrica di campi. In L si consideri l'insieme B di quegli elementi che sono radici di polinomi a coefficienti in A e coefficiente direttore uguale all'unità. Evidentemente $B \supset A$; si può dimostrare che B è un anello. Si chiama chiusura intera dell'anello A nell'estensione L.

Sia ora $\mathbf{Q} \subset K$ un'estensione finita. Si dice anello dei numeri algebrici interi A in K la chiusura intera dell'anello dei numeri interi $\mathbf{Z}$ in K. Se il grado di K su $\mathbf{Q}$ è uguale a n, allora A è uno $\mathbf{Z}$-modulo libero di rango n. In altri termini esiste una base libera $\omega_1, \ldots, \omega_n \in A$: ogni elemento di A si rappresenta univocamente come combinazione lineare $\sum_{i=1}^{n} a_i \omega_i$, $a_i \in \mathbf{Z}$.

Ad esempio se $K = \mathbf{Q}(\sqrt{m})$, con $m \neq 0$ numero intero non divisibile per un quadrato, allora $A = \mathbf{Z} + \mathbf{Z}\sqrt{m}$ per $m \equiv 2, 3 \bmod 4$ e $A = \mathbf{Z} + \mathbf{Z}[(1+\sqrt{m})/2]$ per $m \equiv 1 \bmod 4$. Se $K = \mathbf{Q}(\xi)$, dove ξ è una radice primitiva dell'unità di ordine m, allora $A = \sum_{i=0}^{\varphi(m)-1} \mathbf{Z}\xi^i$, dove $\varphi(m)$ è la funzione di Eulero.

Se L/K è un'estensione algebrica finita di campi, $a \in L$, allora il prodotto per a si può rappresentare con una matrice ad elementi in K rispetto a una base lineare di L su K. Il determinante e la traccia di questa matrice non dipendono dalla scelta della base; si chiamano norma ($\mathrm{N}_{L/K} a$) e traccia ($\mathrm{Tr}_{L/K} a$) dell'elemento a. La norma e la traccia (su $\mathbf{Q}$) di un numero algebrico intero sono usuali numeri interi. Se $\mathbf{Q} \subset K$, $a \in K$, la norma di a è il prodotto di tutti i numeri coniugati con a (immagini di a rispetto alle diverse immersioni isomorfe di K in $\mathbf{C}$) e la traccia di a è la loro somma.

3.2. Ideali.

Sia A l'anello dei numeri algebrici interi di un'estensione finita del campo $\mathbf{Q}$. Ogni ideale $\mathfrak{a}$ non nullo dell'anello A è tale che il quoziente $A/\mathfrak{a}$ è finito; il suo numero di elementi si chiama norma di $\mathfrak{a}$ ($\mathrm{N}(\mathfrak{a})$). Se $(a) = Aa$, $a \in A$, è un ideale principale, allora $\mathrm{N}((a)) = \mathrm{N}(a) = \mathrm{N}_{K/\mathbf{Q}}(a)$.

Se $\mathfrak{p} \subset A$ è un ideale primo non nullo, allora $\mathrm{N}(\mathfrak{p}) = p^f$ (dove p è un numero primo), f è il grado del campo delle classi di resti rispetto al proprio sottocampo primo. Il semigruppo di tutti gli ideali non nulli dell'anello A è generato in maniera libera dagli ideali primi: ogni $\mathfrak{a}$ si rappresenta univocamente nella forma $\mathfrak{a} = \prod_{i=1}^{g} \mathfrak{p}_i^{b_i}$. Se $(p) = \prod_{i=1}^{g} \mathfrak{p}_i^{e_i}$, allora $\sum_{i=1}^{g} e_i f_i = n$, dove $\mathrm{N}(\mathfrak{p}_i) = p_i^{f_i}$ (p numero primo).

Una descrizione esplicita delle leggi delle decomposizioni $(p) = \prod_{i=1}^{g} \mathfrak{p}_i^{e_i}$ o almeno delle loro caratteristiche numeriche g, e_i, f_i in termini del numero primo p è uno dei principali problemi dell'aritmetica dei numeri algebrici. Una sua variante è data dalle leggi relative di decomposizione degli ideali primi dell'anello A, estesi al sopra-anello B, dove $A \subset B$ è l'anello dei numeri algebrici interi di una coppia di estensioni finite del campo $\mathbf{Q}$. L'informazione essenziale sulle leggi di decomposizione delle estensioni abeliane è costituita dalla teoria dei campi di classi (si veda oltre); le leggi non abeliane di estensione sono studiate molto poco.

Se $(p) = \prod_{i=1}^{g} \mathfrak{p}_i^{e_i}$, un ideale $\mathfrak{p}_i$, per il quale $e_i \geq 2$ si dice ramificato (nell'estensione K). Se $e_i \geq 2$ per qualche i, anche il numero primo p si dice ramificato nel campo K. Il numero e_i si dice indice della ramificazione $\mathfrak{p}_i$. Analoghe definizioni si dànno nel caso relativo.

Se $B = A[\alpha]$, dove α è una radice del polinomio irriducibile $f(X) \in A[X]$ con coefficiente direttore 1, e $\mathfrak{p} \subset A$ è un ideale primo, la decomposizione dell'ideale $\mathfrak{p}B$ in B è parallela alla decomposizione $\bar{f}(X) = f(X) \bmod \mathfrak{p} \in A/\mathfrak{p}[X]$. Piú precisamente se $\bar{f}(X) = \prod_{i=1}^{g} \bar{P}_i(X)^{e_i}$ è la decomposizione in fattori irriducibili, allora $\mathfrak{p}B = \prod_{i=1}^{g} \mathfrak{p}_i^{e_i}$, dove $\mathfrak{p}_i = \mathfrak{p}B + P_i(\alpha)B$.

Ad esempio nel campo $\mathbf{Q}(\xi)$, dove ξ è una radice primitiva di ordine m dell'unità, sono ramificati esattamente i divisori di m. Se p non divide m, allora $p = \mathfrak{p}_1, \ldots, \mathfrak{p}_g$, $\mathrm{N}(\mathfrak{p}_i) = p^f$ per ogni i, dove f è un numero intero minimale per il quale $p^f \equiv 1 \bmod m$, $g = \varphi(m)/f$.

Si dice ideale frazionario nel campo K un sotto A-modulo $\mathfrak{a} \subset K$ tale che $c\mathfrak{a} \subset A$ per un opportuno $c \in K^*$. Gli ideali frazionari costituiscono un gruppo, generato in maniera libera dagli ideali $\mathfrak{p}$ dell'anello A.

3.3. Differente e discriminante.

Siano $A \subset B$ gli anelli dei numeri algebrici interi di $K \subset L$. Posto $B' = \{b \in L \mid \mathrm{Tr}_{L/K}(bB) \subset A\}$, si può dimostrare che B' è un ideale frazionario di L. Il suo inverso B'^{-1} si dice differente (relativo) del campo L su K; è un ideale di A. La norma del differente del campo K su $\mathbf{Q}$, moltiplicata per $(-1)^{n(n-1)/2}$, $n = [K : \mathbf{Q}]$, si dice discriminante del campo K; è un numero intero d_K.

Il discriminante del campo K consiste esattamente di quei numeri primi che sono ramificati in quel campo. Il discriminante del campo $\mathbf{Q}(e^{2\pi i/p^r})$ è uguale a $\pm p^{r\varphi(p^r) - p^{r-1}}$.

Il discriminante è il piú semplice e il piú importante invariante del campo dei numeri algebrici. L'unico campo che ha come discriminante l'unità è **Q**; un'altra formulazione dello stesso risultato è la seguente: il campo **Q** non ha estensioni non ramificate non banali. Questo risultato è analogo in teoria dei numeri al fatto che il gruppo fondamentale della sfera di Riemann è banale. Inoltre, esiste soltanto un numero finito di campi con dato discriminante (Minkowski).

Campi diversi da **Q** possono avere estensioni non ramificate ed anche di ordine qualsiasi (Golod-Šafarevič). Esempi di tali campi: $\mathbf{Q}(\sqrt{-30030})$ e $\mathbf{Q}(\sqrt{9699690})$.

Tuttavia l'estensione abeliana massimale non ramificata di un campo K è sempre finita e il suo gruppo di Galois è canonicamente isomorfo al gruppo delle classi di ideali del campo (Hilbert).

3.4. Unità e regolatore.

Le unità del campo K sono gli elementi invertibili del suo anello dei numeri algebrici interi. Esse costituiscono un gruppo abeliano. La sua torsione è finita ed è costituita dalle radici dell'unità che appartengono al campo K. Il suo rango è uguale a $r-1=r_1+r_2-1$, dove r_1 è il numero delle diverse immersioni di K in **R** mentre r_2 è il numero delle diverse coppie di immersioni complesse coniugate di K in **C**. Un invariante piú sottile del gruppo delle unità è il regolatore. Si scelgano $r-1$ immersioni $\sigma_i : K \to \mathbf{C}$ (una per ogni coppia di complessi coniugati) e i generatori $u_1, \ldots, u_r$ del gruppo delle unità rispetto al modulo di torsione. Sia $N_i=1$, se $\sigma_i(K)\subset\mathbf{R}$, $N_i=2$ in caso contrario. Il determinante $\det(N_i \log\sigma_i(u_j))$ non dipende da tale scelta e si dice regolatore del campo K.

3.5. Gruppo di classi e numero di classi.

Tutti gli ideali frazionari del campo K costituiscono un gruppo $\mathcal{J}$, gli ideali frazionari del tipo aA, $a\in K$ costituiscono il suo sottogruppo $\mathcal{J}_0$ degli ideali principali; il quoziente $\mathcal{J}/\mathcal{J}_0$ si dice gruppo delle classi di ideali del campo K. È un gruppo finito; il suo ordine si dice numero di classi del campo K.

3.6. Funzione zeta e relazioni fra il discriminante, il regolatore e il numero di classi.

Ad ogni campo di numeri algebrici K si può associare la sua funzione zeta di Dedekind, che generalizza la costruzione di Eulero del campo **Q**. Precisamente, si ponga $\zeta_K(s)=\sum_{\mathfrak{a}} \frac{1}{\mathrm{N}(\mathfrak{a})^s}$, dove la sommatoria è estesa a tutti gli ideali $\mathfrak{a}$ dell'anello degli interi del campo K. Questa serie converge per ogni numero complesso s con parte reale $\mathrm{Re}\, s>1$ e può essere prolungata in maniera meromorfa

su tutto il piano complesso della variabile s. Nella sua regione di convergenza è valido un analogo del prodotto di Eulero: $\zeta_K(s)=\prod_{\mathfrak{p}}(1-\mathrm{N}(\mathfrak{p})^{-s})^{-1}$, dove $\mathfrak{p}$ è un qualunque ideale primo dell'anello A.

La funzione zeta contiene molte informazioni aritmetiche sul campo K. Ci si limita qui a formulare il seguente, classico, risultato: nel punto $s=1$ la funzione $\zeta_K(s)$ ha un polo del primo ordine e in esso il residuo è uguale a $(2^{r_1+r_2}\pi^{r_2}Rh)/m\sqrt{|d|}$ dove R è il regolatore, h il numero di classi, d il discriminante e m il grado del campo.

Questa formula conduce a conseguenze particolarmente importanti nei casi in cui il residuo di $\zeta_K(s)$ si può calcolare immediatamente, senza rivolgersi agli invarianti aritmetici del campo K. Poiché m e d, nei casi concreti, si calcolano senza difficoltà, la formula costituisce una relazione fra R e h.

Ad esempio sia K l'estensione quadratica del campo **Q** con discriminante D. Ad essa è naturale porre in corrispondenza un carattere quadratico rispetto al modulo $|D|$, vale a dire l'omomorfismo di gruppi $\chi : (\mathbf{Z}/|D|)^* \to (\mp 1)$. Se $D<0$, vale a dire K è un campo quadratico complesso, allora $r_1=r_2=0$, $R=1$, e un calcolo diretto del residuo della funzione zeta conduce a una formula per il numero di classi del tipo $h=-\frac{1}{|D|}\sum_{(x,D)=1}\chi(x)x$. Se $D>0$, allora il rango del gruppo delle unità è uguale a 1 e scegliendo il suo generatore rispetto al modulo di torsione («unità principale») $\varepsilon>1$, si ottiene un'identità del tipo

$$h\log\varepsilon=-\sum_{\substack{(x,D)=1\\ 0<x<(D/2)}}\chi(x)\log\sin\frac{\pi x}{D}.$$

Quest'ultima formula può essere reinterpretata come segue: il numero

$$\eta=\frac{\pi\sin\frac{\pi x}{D}}{\pi\sin\frac{\pi y}{D}} \qquad 0<x,\ y<\frac{D}{2},\ \chi(x)=-1,\ \chi(y)=+1$$

come si può verificare, è un'unità del campo K e $\varepsilon=\eta^h$. In altri termini, a meno di un fattore inessenziale, il numero di classi di un campo quadratico reale coincide con l'ordine del quoziente di tutte le sue unità rispetto al sottogruppo generato dalle unità costruibili η.

Un analogo risultato ha luogo per il campo $K=\mathbf{Q}(\xi+\xi^{-1})$, dove ξ è una radice primitiva d'ordine primo l dell'unità. Qui i $\vartheta_a=\sin a\pi/l\cdot(\sin\pi/l)^{-1}$ (numeri) sono unità reali positive costruibili del campo e il numero di classi è uguale all'indice del sottogruppo da esse generato nel gruppo di tutte le unità reali positive del campo.

Questi risultati sorprendono già per il fatto di stabilire un'uguaglianza fra gli ordini di due gruppi (il gruppo di classi e il quoziente delle unità), senza che i due gruppi siano isomorfi (essi in effetti non sono sempre isomorfi, come mostrano le tabelle). A volte tuttavia, utilizzando l'azione del gruppo di Galois del campo K, si può stabilire un isomorfismo fra opportune componenti di

questi gruppi e ciò conduce a risultati molto avanzati. Questi problemi costituiscono l'oggetto della moderna teoria di Iwasawa e di ulteriori, a volte ipotetiche, generalizzazioni.

3.7. Gruppo di Galois e aritmetica.

La piú semplice informazione sui legami del gruppo di Galois con gli ideali primi è fornita dal seguente risultato. Sia $K \subset L$ un'estensione finita di Galois con gruppo G, $A \subset B$ i corrispondenti anelli dei numeri interi, $\mathfrak{y} \subset A$ un ideale primo. Allora $\mathfrak{y}B = (\mathfrak{p}_1 \ldots \mathfrak{p}_g)^l$, dove $\mathfrak{p}_i$ sono ideali primi sui quali il gruppo G agisce transitivamente. In particolare i sottogruppi stazionari $G_{\mathfrak{p}_i}$ degli ideali $\mathfrak{p}_i$ sono a due a due coniugati. Il gruppo $G_{\mathfrak{p}_i}$ agisce naturalmente sulla estensione dei campi finiti $A/\mathfrak{y} \subset B/\mathfrak{p}_i$. Sia $I_{\mathfrak{p}_i} \subset G_{\mathfrak{p}_i}$ il sottogruppo normale degli elementi che vi agiscono in maniera banale. Allora il quoziente $G_{\mathfrak{p}_i}/I_{\mathfrak{p}_i}$ è isomorfo al gruppo completo di Galois dell'estensione $A/\mathfrak{y}_i \subset B/\mathfrak{p}_i$. Quest'ultima è generata dall'automorfismo (relativo) di Frobenius $\Phi_{\mathfrak{p}_i} : x \mapsto x^{N(\mathfrak{y}_i)}$. Ciò permette di considerare $\Phi_{\mathfrak{p}_i}$ come un elemento del gruppo $G_{\mathfrak{p}_i}/I_{\mathfrak{p}_i}$. Se l'ideale $\mathfrak{y}$ non è ramificato nel campo L, allora tutti i sottogruppi $I_{\mathfrak{p}_i}$ sono banali e si ottiene una classe di elementi coniugati $\Phi\mathfrak{y} = (\text{classe } \Phi_{\mathfrak{p}_i})$ nel gruppo $G = \mathrm{Gal}(L/K)$. In tal modo a ogni ideale non ramificato in L del campo K si associa la sua classe di Frobenius $\Phi_{\mathfrak{y}} \subset G$. Se il gruppo di Galois G è abeliano, i $\Phi_{\mathfrak{y}}$ sono semplicemente gli elementi di G. In questo caso $\Phi\mathfrak{y}$ si denota spesso $(\mathfrak{y}, L/K)$ e si dice simbolo di Artin. Si indica con $I(\mathfrak{D})$ il gruppo degli ideali frazionari del campo K, mutuamente primi, con differente $\mathfrak{D}$, dell'estensione L/K e si estende il simbolo di Artin a un omomorfismo di Artin dei gruppi $I(\mathfrak{D}) \to G$ rispetto al prodotto: $\mathfrak{a} \mapsto (\mathfrak{a}, L/K)$, dove $(\pi\mathfrak{y}_i^{n_i}, L/K) = \pi(\mathfrak{y}_i, L/K)^{n_i}$.

3.8. Teoria dei campi di classi.

Il risultato principale della teoria dei campi di classi è la dimostrazione che l'omomorfismo di Artin $I(\mathfrak{D}) \to G$ è suriettivo e la descrizione del suo nucleo in termini del campo L. Insieme alle proprietà funtoriali di questo omomorfismo (rispetto a L, ma anche rispetto a K), questo risultato permette di dare una descrizione, sorprendente per bellezza e completezza, del gruppo di Galois dell'estensione abeliana massimale $G^{ab} = \mathrm{Gal}(K^{ab}/K)$ d'ogni campo di numeri algebrici K. Senza soffermarsi sui dettagli, si formulerà qui il risultato conclusivo.

Si consideri il gruppo topologico $(\mathbf{R}^*)^{r_1} \times (\mathbf{C}^*)^{r_2} \times \prod_{\mathfrak{y}} K_{\mathfrak{y}}^*$ dove $\mathfrak{y}$ varia negli ideali primi del campo K, $K_{\mathfrak{y}}^*$ è il completamento $\mathfrak{y}$-adico di K. Si denoti con $\mathfrak{J}_K$ il sottogruppo costituito dai vettori infiniti $(a_\infty \in \mathbf{R}^{*r_1} \times \mathbf{C}^{*r_2}, a_{\mathfrak{y}} \in K_{\mathfrak{y}}^*)$ tali che $a_{\mathfrak{y}}$ sia un'unità $\mathfrak{y}$-adica per ogni $\mathfrak{y}$, eccetto un numero finito. Il gruppo $\mathfrak{J}_K$ si dice gruppo degli adeli del campo K. Si ha l'immersione canonica $K^* \to \mathfrak{J}_K$: su ogni componente essa corrisponde all'immersione del campo K nel suo completamento topologico $\mathbf{R}$, $\mathbf{C}$ oppure $K_{\mathfrak{y}}$ rispetto a una metrica. Il quoziente $C_K = \mathfrak{J}_K/K$ si dice gruppo di classi degli adeli del campo K. Come gruppo

topologico non è limite proiettivo di gruppi finiti, come deve essere un gruppo topologizzato secondo Krull: ciò è dovuto alla presenza di una componente connessa dell'unità $C_K^0 \subset C_K$ che proviene dal sottogruppo $\mathbf{R}^{*r_1} \times \mathbf{C}^{*r_2} \subset \mathfrak{J}_K$. Il quoziente rispetto ad essa C_K/C_K^0 è già profinito e il teorema fondamentale della teoria dei campi di classi è costituito dalle seguenti affermazioni:

a) Tutti gli omomorfismi di Artin si «incollano» in un omomorfismo $\mathfrak{J}_K \to \mathrm{Gal}(K^{ab}/K)$; questo omomorfismo è suriettivo.
b) Questo omomorfismo induce un isomorfismo $C_K/C_K^0 \to \mathrm{Gal}(K^{ab}/K) = G^{ab}$.
c) Il sottogruppo $\mathfrak{J}_K$ corrispondente agli elementi di G^{ab} che lasciano fissa un'estensione abeliana finita L/K, è uguale a $K^* \mathrm{N}_{L/K} \mathfrak{J}_L$ (non ci si soffermerà a definire l'applicazione $\mathrm{N}_{L/K} : \mathfrak{J}_L \to L_K$, abbastanza evidente).

Tutti gli oggetti che figurano in questa descrizione possono essere calcolati molto chiaramente nel caso $K = \mathbf{Q}$, grazie al classico teorema di Kronekker-Weber per cui il campo $\mathbf{Q}^{ab}$ coincide col campo generato su $\mathbf{Q}$ da tutte le radici dell'unità. Da qui segue l'isomorfismo canonico $G^{ab} = \varprojlim_m (\mathbf{Z}/m\mathbf{Z})^*$, limite rispetto al sistema proiettivo degli interi m con la relazione d'ordine $m_1 > m_2$ se m_1 è divisibile per m_2, che determina un omomorfismo di restrizione rispetto al modulo $m_2 : \mathbf{Z}/m_1\mathbf{Z} \to \mathbf{Z}/m_2\mathbf{Z}$. Questo limite è isomorfo a $\prod_p \mathbf{Z}_p^*$ ($\mathbf{Z}_p^*$ è il gruppo delle unità p-adiche), e non è difficile identificare quest'ultimo gruppo con $C_{\mathbf{Q}}/C_{\mathbf{Q}}^0$.

Nel caso di un campo generale K, le dimostrazioni della teoria dei campi di classi sono eccessivamente indirette e basate, grosso modo, sul confronto degli ordini dei gruppi quoziente finiti $\mathbf{C}_K/N_{L/K}(\mathbf{C}_L)$ e $\mathrm{Gal}(L/K)$ per le variabili K e L. La presenza dell'omomorfismo di Artin e di disuguaglianze fra questi ordini, da diversi lati permette alla fine di stabilire l'isomorfismo cercato. Le valutazioni degli ordini nei gruppi di adeli sono basate su calcoli $\mathfrak{y}$-adici locali, nei gruppi di Galois in esplicite costruzioni di estensioni abeliane per mezzo dell'aggiunta delle radici dell'unità e quindi di radici di elementi del campo. In effetti un metodo banale per costruire un campo, che evidentemente contiene K^{ab}, consiste nell'aggiungere a $K\mathbf{Q}^{ab} = \bigcup_m K(e^{2\pi i/m})$ tutte le radici degli elementi di $K\mathbf{Q}^{ab}$. Tuttavia questa estensione non è abeliana e la descrizione della sua sottoestensione abeliana massimale, o almeno del gruppo di Galois di quest'ultima, non è affatto banale. Il secondo problema, come si è visto, è già stato risolto, ma il primo rimane, insieme alle generalizzazioni non abeliane della teoria dei campi di classi, uno dei problemi fondamentali dell'aritmetica dei numeri algebrici.

3.9. Teoria locale dei campi di classi.

In tutta l'esposizione precedente si è rimasti su un punto di vista globale; il risultato del § 3.8 conduce a pensare che la considerazione di completamenti locali del campo $K_{\mathfrak{y}}$, del gruppo dei suoi adeli $\mathfrak{J}_K$, ecc. fin dall'inizio si sarebbe

rivelata piú utile. In effetti questo punto di vista è oggi quello principale. La sua realizzazione successiva è legata in primo luogo ai nomi di Chevalley, Artin e Whaples, Tate, Weil, Eichler, Ono, Tamagawi. In questo punto ci si limita a formulare il teorema fondamentale della teoria locale dei campi di classi per i campi del tipo $K_{\mathfrak{y}}$. Esso afferma che esiste un isomorfismo canonico fra il gruppo $\mathrm{Gal}(K_{\mathfrak{y}}^{ab}/K_{\mathfrak{y}})$ e il completamento del gruppo $K_{\mathfrak{y}}^{*}$ nella topologia di Krull. A «livello finito» il gruppo di Galois di un'estensione abeliana finita $K_{\mathfrak{y}} \subset L_{\mathfrak{y}}$ è canonicamente isomorfo al gruppo quoziente $K_{\mathfrak{y}}^{*}/N_{L_{\mathfrak{y}}/K_{\mathfrak{y}}}(L_{\mathfrak{y}}^{*})$.

3.10. Il problema di una costruzione esplicita delle estensioni abeliane.

Oltre al teorema di Kronecker-Weber, questo problema è risolto per le estensioni quadratiche immaginarie K del campo $\mathbf{Q}$ nella classica teoria del prodotto complesso di funzioni ellittiche, che rimane un modello per ulteriori ricerche.

L'idea della costruzione consiste nel considerare le radici dell'unità come punti d'ordine finito su un gruppo algebrico moltiplicativo G_m e quindi sostituire G_m con un opportuno gruppo algebrico X_K, legato al campo K, nella speranza che l'aggiunta a K dei suoi punti d'ordine finito fornisca K^{ab} o per lo meno una sottoestensione abbastanza grande di K^{ab}.

Per i campi quadratici immaginari K quest'idea si realizza nel modo seguente. L'anello dei numeri interi A di questo campo reticola il piano complesso $\mathbf{C}$. L'insieme delle funzioni meromorfe su $\mathbf{C}$, periodiche rispetto a questo reticolo, costituisce il campo delle funzioni ellittiche su una riemanniana algebrica. La curva algebrica corrispondente può essere data dall'equazione di Weierstrass $y^2 = 4x^3 - g_2 x - g_3$. Il suo invariante $j = 1728 g_2^3 (g_2^3 - 27 g_3^2)^{-1}$ è un numero algebrico intero determinato a meno di coniugio sul campo K. Il primo risultato della teoria del prodotto complesso consiste nel fatto che $\tilde{K} = K(j)$ è un'estensione abeliana non ramificata massimale del campo K, cioè il cosiddetto campo di classi di Hilbert. Dal teorema fondamentale dei campi di classi (§ 3.9) si deduce che il suo gruppo di Galois è isomorfo al gruppo delle classi di ideali del campo K. Si aggiungano ora a $\tilde{K}$ le coordinate di tutti i punti d'ordine finito sulla curva ellittica con invariante j descritta sopra. Il risultato coincide col campo K^{ab}. L'azione del gruppo di Galois G^{ab}, rappresentata dai simboli di Artin, su j e sulle coordinate dei punti d'ordine finito, si può pure descrivere in maniera completamente esplicita, in analogia alla descrizione dell'azione di $\mathrm{Gal}(\mathbf{Q}^{ab}/\mathbf{Q})$ sulle radici dell'unità.

Recentemente Lubin e Tate hanno dimostrato che un analogo di questa costruzione sussiste per ogni campo locale K. A questo scopo occorre sostituire la curva ellittica con un gruppo formale. Sia A l'anello dei numeri interi del campo K. Un gruppo formale (piú precisamente una legge gruppale formale) su A è una serie formale di potenze $F(X,Y) \in A[\![X,Y]\!]$ con le seguenti proprietà: *a)* $F(X,F(Y,Z)) = F(F(X,Y),Z)$, *b)* $F(X,Y) = F(Y,X)$, $F(X,0) = X$. Sia $\mathfrak{m} \subset A$ un ideale massimale. Per mezzo di F si può introdurre in esso una legge di composizione $*$: $x * y = F(x,y)$. Rispetto a questa legge di composizione si trasforma in un gruppo abeliano (in effetti è addirittura un gruppo di Lie su K).

Ad esempio se $F(X,Y) = X + Y + XY$, questo gruppo è soltanto $(1+\mathfrak{m})^*$.

Si dice endomorfismo di una legge formale F una serie $f \in A[\![X]\!]$ tale che $f(F(X,Y)) = F(f(X), f(Y))$. Un'applicazione $f: \mathfrak{m} \to \mathfrak{m}$ definisce in questo caso un endomorfismo del gruppo $(\mathfrak{m}, *)$.

Sia $q = \mathrm{Card}\, A/\mathfrak{m}$, π un generatore dell'ideale $\mathfrak{m}$. Esiste allora un'unica legge gruppale F_π per la quale $f(X) = \pi X + X^2$ sia un endomorfismo. Si tratta dell'endomorfismo di prodotto per π.

Si denoti con L l'estensione del campo K generata *a)* da tutte le radici dell'unità di grado primo con q; *b)* da tutti i punti π^n secondo l'ordine del gruppo definito dalla legge F_π, vale a dire dalle radici dei polinomi $f(f \ldots (f(X)) \ldots)$. Allora L è l'estensione abeliana massimale del campo locale K. L'azione di K^* su essa si può descrivere esplicitamente grazie all'isomorfismo di $\mathrm{Gal}(L/K)$ con $\hat{K}^*$, come nella teoria classica del prodotto complesso.

3.11. Ulteriori idee.

A causa delle dimensioni limitate di questo articolo si sono dovute lasciar da parte molte idee che si trovano al centro dell'attenzione degli studiosi moderni. Fra queste rientrano, in particolare, i problemi di calcolo della coomologia dei gruppi di Galois di campi globali e locali; la descrizione delle rappresentazioni dei gruppi di Galois, in particolare sulle varietà algebriche di coomologia l-adica; le connessioni di queste ultime con le funzioni zeta e con le forme automorfe e il «programma di Langlands» di costruire una teoria non commutativa dei campi di classi; il ruolo della K-teoria algebrica nell'aritmetica contemporanea [si vedano a questo riguardo Langlands 1974; Tate 1974].

4. *Numeri trascendenti.*

4.1. Il problema della trascendenza o dell'indipendenza algebrica.

In questo paragrafo si darà un cenno sulla dimostrazione del teorema da cui deriva la trascendenza dei numeri e, π e α^β ($\alpha \neq 0, 1$ numero algebrico, β numero irrazionale algebrico).

Il problema della trascendenza di α^β fu posto da Eulero e incluso da Hilbert nel suo famoso elenco di problemi (1900). Esso fu risolto indipendentemente da Gel'fond e da Schneider (1934) dopo lavori anticipatori di Gel'fond del 1929. Qui viene seguito Lang [1966], il quale ha semplificato e generalizzato la dimostrazione di Gel'fond e Schneider. In conclusione è riportato un elenco del tutto incompleto di altri risultati e problemi.

L'idea comune alla maggior parte dei metodi dimostrativi della trascendenza o dell'indipendenza algebrica di numeri concreti e di funzioni è legata alla loro rappresentazione sotto forma di valori di funzioni analitiche, la cui

trascendenza e indipendenza algebrica sono incomparabilmente piú semplici da determinare. Dopo di ciò si costruiscono alcune funzioni analitiche ausiliarie: i loro valori sono maggiorati con considerazioni analitiche e si dimostra che queste maggiorazioni contraddicono valutazioni aritmetiche dal basso, che si ottengono con considerazioni aritmetiche, se i numeri originari sono algebrici. Le valutazioni dal basso, in ultima analisi, sono legate semplicemente al fatto che un numero intero non nullo ha sempre modulo ≥ 1.

Una funzione algebrica (eventualmente plurivoca) $f(z)$ della variabile complessa z possiede le seguenti proprietà:

a) per prosecuzione analitica, f assume un numero limitato di valori in ogni punto, vale a dire ha un numero finito di fogli.
b) f assume ogni valore prefissato in un numero limitato di punti.
c) $\max_{|z|=R} |f(z)| < R^n$ per $R > R_0$ e n opportuno.
d) f non ha singolarità essenziali.

Pertanto, per dimostrare la trascendenza della funzione $f(z)$ è sufficiente verificare che non vale una delle proprietà elencate. Ad esempio e^z è trascendente poiché $e^z = 1$ per ogni $z \in 2\pi i\mathbf{Z}$ (non vale la *b*)), e anche poiché e^R cresce piú velocemente di qualsiasi potenza di R (non vale la *c*)). La funzione $\log z$ è trascendente perché non vale la condizione *a*).

Le funzioni $e^{\alpha_1 z}, \ldots, e^{\alpha_n z}$ sono algebricamente indipendenti, se $\alpha_1, \ldots, \alpha_n$ sono numeri linearmente indipendenti su $\mathbf{Q}$. In effetti, considerando per semplicità che α_i sia reale, si supponga che sussista una relazione polinomiale fra le $e^{\alpha_i z}$. Sarà della forma $\Sigma a_{m_1 \ldots m_n} e^{(m_1\alpha_1 + \ldots + m_n\alpha_n)z} = 0$. Dall'indipendenza lineare su $\mathbf{Q}$ degli α_i segue che tutti i numeri $\sum_{j=1}^{n} m_j\alpha_j$ sono diversi e non uguali a zero per $(m_j) \neq (0)$. Si scelga fra questi il massimo, sia ν. L'elemento $a_{m_1 \ldots m_\nu} e^{\nu z}$ sarà dominante nella somma per $z \in \mathbf{R}$, $\nu z \to \infty$ e non può essere ridotto con gli elementi rimanenti: contraddizione.

La maggior parte delle dimostrazioni necessarie per le applicazioni alla indipendenza trascendente di funzioni non sono essenzialmente piú complicate di questa, benché i dettagli tecnici possano complicarsi.

4.2. Funzioni con ordine di crescenza finito.

Per la formulazione del teorema sulla trascendenza occorrono i seguenti concetti. Una funzione intera è una funzione analitica (univoca) $f : \mathbf{C} \to \mathbf{C}$. Una funzione intera f ha ordine di crescenza $\leq \rho$ se esiste una costante $C > 1$ tale che per $|z| \leq R$ sia verificata la valutazione $|f(z)| \leq C^{R^\rho}$ per ogni R abbastanza grande. Una funzione meromorfa di ordine $\leq \rho$ è il rapporto di due funzioni intere di ordine $\leq \rho$.

4.3. Formulazione del teorema e sue conseguenze.

$f_1, \ldots, f_N$ siano funzioni meromorfe con ordine di crescenza finito, comprendenti due funzioni algebricamente indipendenti. Si supponga che le f_i soddisfino un sistema di equazioni differenziali del tipo $\partial f_i / \partial z = P_i(f_1, \ldots, f_N)$, dove P_i è un polinomio a coefficienti algebrici, $i = 1, \ldots, N$. Allora il numero dei punti $w_i \in \mathbf{C}$ (esclusi i poli delle funzioni f_i) nei quali tutte le funzioni f_i assumono valori in un dato campo K, $[K : \mathbf{Q}] < \infty$, è finito.

Supponendo $f_1 = z$, $f_2 = e^z$, si trova che il numero e^α è trascendente se $\alpha \neq 0$ è algebrico. In effetti, altrimenti, dal campo $K = \mathbf{Q}(\alpha, e^\alpha)$ si otterrebbe che f_1 e f_2 assumono valori di K in un insieme infinito di punti α, 2α, 3α, ... In particolare è trascendente il numero π, poiché $e^{2\pi i} = 1$ è algebrico.

Supponendo $f_1 = e^{\beta t}$, $f_2 = e^t$, dove β è un numero algebrico irrazionale, si ha che $\alpha^\beta = e^{\beta \log \alpha}$ (dove $\alpha \neq 0, 1$ è algebrico) è un numero trascendente: altrimenti f_1 e f_2 assumerebbero valori algebrici in un campo di grado finito in tutti i punti $m \log \alpha$, $m \geq 1$. L'indipendenza algebrica di f_1 e f_2 segue dall'irrazionalità di β, vale a dire dall'indipendenza lineare fra 1 e β su $\mathbf{Q}$.

Si osservi ancora che il numero dei punti in cui le f_i possono assumere valori di un dato campo K, si può valutare esplicitamente nelle ipotesi del teorema: se K contiene i coefficienti dei polinomi P_i, e gli ordini di f_i non superano ρ, questo numero è $\leq 20\rho[K : \mathbf{Q}]$.

4.4. Componenti aritmetiche della dimostrazione.

La dimostrazione è basata sulla costruzione di funzioni analitiche ausiliarie, soggette a condizioni lineari e connesse all'insieme di punti nei quali le f_i assumono valori algebrici.

Nella costruzione delle funzioni ausiliarie svolgono un ruolo importante i lemmi sulla soluzione delle equazioni lineari, applicati per la prima volta da Siegel ad analoghe questioni.

a) *Valutazioni su* $\mathbf{Q}$. Sia (a_{ij}) una matrice intera con r righe e n colonne, $n > r$, $|a_{ij}| < B$. Si consideri il sistema di equazioni lineari $\sum_{j=1}^{n} a_{ij}x_j = 0$: esso ha soluzione intera non nulla con $|x_j| \leq 2(nB)^{r/(n-r)}$, $j = 1, \ldots, n$.

La dimostrazione procede nel modo seguente: il numero dei vettori (y_i) con $y_i = \sum_{j=1}^{n} a_{ij}x_j$, $|x_j| \leq (nB)^{r/(n-r)}$, è maggiore del numero dei vettori tali che $|y_i| \leq \leq \sum_{j=1}^{n} |a_{ij}||x_j| \leq nB(nB)^{r/(n-r)}$. Pertanto ogni coppia di vettori distinti (x_j), (x'_j) fornisce lo stesso valore (y_i) e la loro differenza dà una soluzione non nulla del nostro sistema di equazioni.

b) *Valutazioni su* K, $[K : \mathbf{Q}] < \infty$. Si consideri lo stesso problema del punto *a*), però con $a_{ij} \in A$, essendo A l'anello dei numeri interi del campo K.

Se $S \subset K$, poniamo $\|S\| = \max |gs|$, $s \in S$, $g \in \{$immersioni di K in $\mathbf{C}\}$. Se $\|(a_{ij})\| \leq \leq B$ esiste una soluzione del sistema di equazioni del punto *a*) con $x_j \in A$ e $\|x_j\| \leq C_1(C_2 nB)^{r/(n-r)}$, dove C_1 e C_2 sono opportune costanti, dipendenti solo da K. Per la dimostrazione occorre decomporre a_{ij}, x_j rispetto a una base fissata di A su $\mathbf{Z}$ e applicare il risultato precedente insieme a evidenti valutazioni di $\|S\|$ per mezzo di $|S|$.

c) *Valutazione di* $\|P\|$, *dove* P *è un polinomio nelle* f_i. Sia $S \subset K$; se $B \in \mathbf{Z}$, $B > 0$ è un numero intero tale che $BS \subset A$, si dice che B è il denominatore di S e si denota ogni denominatore di S con $\operatorname{den} S$. $f_1, \ldots, f_N$ siano funzioni olomorfe nell'intorno di un punto $w \in \mathbf{C}$ e $f_i(w) \in K$, $i = 1, \ldots, N$; sia $\partial f_i / \partial z = P_i(f_1, \ldots, f_N)$, $P_i \in K[T_1, \ldots, T_N]$. Sia infine $P \in K[T_1, \ldots, T_N]$ un polinomio di grado $\leq r$, $f = P(f_1, \ldots, f_N)$, $\|P\| = \|$(coefficienti di P)$\|$. Allora per ogni intero $K > 0$ valgono, con una opportuna costante C_1, le valutazioni seguenti:

$$\left\| \frac{d^k}{dz^k} f(w) \right\| \leq \|P\| r^k k! C_1^{k+r}$$

$$\min \operatorname{den} \frac{d^k}{dz^k} f(w) \leq \operatorname{den}(P) \cdot C_1^{k+r}$$

Questo risultato si dimostra per induzione non complicata rispetto a k.

4.5. Costruzione delle funzioni ausiliarie.

Sotto le ipotesi del § 4.2 si scelgano m punti w_j nei quali $f_i(w_j) \in K$ (i coefficienti dei $P_i \in K$). Una maggiorazione di m si ottiene da valutazioni opposte per i valori della funzione ausiliaria e delle sue derivate nei punti w_j. Si scelga come F il polinomio $F = \sum_{i,j=1}^{r} b_{ij} f^i g^j$, dove $f, g \in \{f_1, \ldots, f_N\}$ sono algebricamente indipendenti. Come r si prenda un numero divisibile per $2m$; piú avanti lo si renderà molto grande. S'impongano inoltre le seguenti condizioni ai coefficienti di F: *a*) $b_{ij} \in A$; *b*) $\|b_{ij}\| \leq Cn^{2n}$; *c*) $\frac{d^k}{dz^k} F(w_j) = 0$ per $1 \leq j \leq m$, $1 \leq \leq k \leq n$. Le valutazioni del § 4.2 permettono di mostrare l'esistenza di $(b_{ij}) \neq (0)$ con queste proprietà, se $n = r^2/2m$ per un'opportuna costante C. Per essi $F \not\equiv 0$, giacché f e g sono algebricamente indipendenti.

4.6. Valutazione dei valori delle derivate di F.

Sia $\left(\frac{d^s F}{dz^s}(w_j)\right) \neq (0)$ ma $\left(\frac{d^l F}{dz^l}(w_j)\right) = (0)$ per $l \leq s-1$. Allora $s \geq n$ e si può ritenere che $\gamma = \frac{d^s}{dz^s} F(w_j) \neq 0$. Il numero s tende a ∞ insieme a r. In forza del § 4.4, *c*), esiste un numero intero $c \in \mathbf{Z}$, $0 < c < CC_1^s$, tale che $c\gamma \in A$ e $\|\gamma\| < Cs^{5s}$. Allora $1 \leq |N_{K/\mathbf{Q}}(c\gamma)| \leq Cs^{5([K:\mathbf{Q}]-1)s} |\gamma|$.

D'altro lato esiste una funzione intera ϑ di ordine $\leq \rho$ con $\vartheta(w_1) \neq 0$ e ϑf, ϑg intere. Allora $\vartheta^{2r} F$ è una funzione intera e $G(z) = \vartheta(z)^{2r} \cdot F \cdot \prod_{\nu=1}^{m} (z - w_\nu)^s$ è una funzione intera. Il valore $G(w_1)$ differisce da $\gamma = \frac{d^s}{dz^s} F(w_1)$ per un fattore limitato dalla quantità $C_1^s s!$ D'altro lato $|G(w_1)| \leq \max_{|z|=R} |H(z)|$ se $R \geq |w_1|$. Prendendo n abbastanza grande, ad esempio s, e quindi ponendo $R = s^{1/2\rho}$, si ottiene senza difficoltà la valutazione $C_2^s \cdot s^{s[4-(m/2\rho)]}$ per $\max_{|z|=R} |H(z)|$.

Confrontando la maggiorazione che ne deriva per ν con la valutazione dal basso al termine del precedente paragrafo si ottiene una contraddizione se $m > 20\rho[K : \mathbf{Q}]$; ciò dimostra il teorema.

4.7. Ulteriori risultati e problemi.

a) Il teorema di Gel'fond e Schneider sulla trascendenza di α^β può essere riformulato come segue: se α, β sono numeri algebrici non nulli e $\log\alpha$, $\log\beta$ sono linearmente indipendenti su $\mathbf{Q}$, allora sono linearmente indipendenti su $\bar{\mathbf{Q}}$. Alla fine degli anni '60 Baker, sviluppando il metodo di Gel'fond e Schneider, mostrò, per ogni $n \geq 2$, che se $\alpha_1, \ldots, \alpha_n$ sono numeri algebrici non nulli e i loro logaritmi sono linearmente indipendenti su $\mathbf{Q}$, allora sono linearmente indipendenti su $\bar{\mathbf{Q}}$.

Un analogo risultato, con l'esponenziale al posto del logaritmo, era stato stabilito da Weierstrass nel 1855.

b) Esistono numerosi risultati relativi all'approssimazione di numeri trascendenti noti per mezzo di numeri algebrici.

Sia ξ un numero algebrico; si denoti con $H(\xi)$ il massimo modulo dei coefficienti di un polinomio irriducibile di cui ξ sia una radice. Allora per ogni $\varepsilon > 0$ vale la valutazione

$$\left| \frac{\log\alpha}{\log\beta} - \xi \right| > \exp(-|\log H|^{2+\varepsilon})$$

per $H > H_0(\alpha, \beta, \varepsilon, [\mathbf{Q}(\varepsilon) : \mathbf{Q}])$ (Gel'fond, 1949). Baker e Feldman hanno esteso queste valutazioni alle forme lineari di piú logaritmi.

c) Molta attenzione è stata posta al problema della trascendenza dei valori di funzioni speciali, legate all'uniformizzazione delle varietà algebriche. Ad esempio Schneider ha dimostrato che data la curva ellittica $y^2 = x^3 - g_2 x - g_3$, g_2 e g_3 algebrici, i punti algebrici P su di essa hanno argomenti trascendenti di Weierstrass $\int_\infty^P (dx/y)$. Analoghi risultati valgono per le funzioni di Bessel, ecc.

d) Schanuel ha formulato la seguente ipotesi: se i numeri $\alpha_1, \ldots, \alpha_n$ sono linearmente indipendenti su $\bar{\mathbf{Q}}$, fra i numeri $\alpha_1, \ldots, \alpha_n, e^{\alpha_1}, \ldots, e^{\alpha_n}$ ve ne sono un numero $\geq n$ algebricamente indipendenti su $\mathbf{Q}$. Recentemente Čudnovskij ha mostrato che per insiemi opportuni $(\alpha_1, \ldots, \alpha_n)$, $n \to \infty$, la quantità di numeri algebricamente indipendenti fra $(\alpha_1, \ldots, \alpha_n; e^{\alpha_1}, \ldots, e^{\alpha_n})$ tende all'infinito. [JU. I. M.].

Bourbaki, N.
1960 *Eléments d'histoire des mathématiques*, Hermann, Paris (trad. it. Feltrinelli, Milano 1963).

Lang, S.
1966 *Introduction to Transcendental Numbers*, Addison-Wesley, Reading Mass.

Langlands, R. P.
[1974] *Some Contemporary Problems with Origins in the Jugendtraum (Hilbert's Problem 12)*, in F. E. Browder (a cura di), *Mathematical Developments Arising from Hilbert Problems. Symposium in Pure Mathematics, Northern Illinois University, 1974*, American Mathematical Society, Providence R.I. 1976, pp. 401-18.

Šafarevič, I. R.
1962 *Polja algebraičeskich čisel*, in *Proceedings of the International Congress of Mathematicians*, Institut Mittag-Leffler, Djursholm, pp. 163-76.

Tate, J.
[1974] *Problem 9: the General Reciprocity Law*, in F. E. Browder (a cura di), *Mathematical Developments Arising from Hilbert Problems. Symposium in Pure Mathematics, Northern Illinois University, 1974*, American Mathematical Society, Providence R.I. 1976, pp. 311-22.

I termini si riferiscono in primo luogo a numeri e a funzioni (cfr. **numero, funzione, divisibilità**) e forniscono una classificazione legata, in un certo senso, alla loro complessità (cfr. **semplice/complesso**). A partire dal concetto di numero e di figura geometrica (cfr. **geometria e topologia, curve e superfici, matematiche**) la storia di questa classificazione (cfr. **sistematica e classificazione**) si sviluppa attorno alle idee fondamentali di linguaggio algebrico per rappresentare numeri e **operazioni** (cfr. **applicazioni, rappresentazione**), di **interpretazione** dei numeri come **misura**, di loro applicazione ai problemi dell'**infinito** (cfr. anche **continuo/discreto, insieme**). Oggi costituiscono un capitolo importante dell'algebra (in cui si riferiscono a elementi delle principali **strutture matematiche**) e della teoria (cfr. **teoria/modello**) dei numeri, in cui s'impongono problemi di dimostrazione della trascendenza di varie classi di numeri (cfr. anche **deduzione/prova, dipendenza/indipendenza**) o di **approssimazione** mediante numeri di classe inferiore.

Simmetria

La prima parte di questo articolo è costituita da osservazioni sulla storia del principio di simmetria; essa contiene anche una descrizione delle sue applicazioni alla fisica moderna. Nella seconda parte vengono esposte in maniera frammentaria nozioni introduttive sul principale apparato matematico della teoria della simmetria: i gruppi e le loro rappresentazioni. La materia è talmente ampia che perfino il semplice tentativo di denominare i temi che le sono collegati non è possibile; si veda anche l'articolo «Dualità» in questa stessa *Enciclopedia* che sostanzialmente è dedicato a un aspetto della simmetria.

1. *Nozioni generali.*

1.1. Osservazioni storiche.

La parola greca σύμμετρος denota 'commensurato, proporzionale'. Euclide utilizza questa parola in relazione ai segmenti commensurabili, cioè i segmenti il rapporto delle cui lunghezze è razionale (cfr. l'articolo «Razionale/algebrico/trascendente» in questa stessa *Enciclopedia*).

Hermann Weyl, all'inizio della prima lezione del suo *Symmetry* [1952], scrive che «la parola tedesca *Ebenmass* è un ottimo equivalente della parola greca "simmetria": entrambe esprimono anche l'idea di "misura mediana", il giusto mezzo al quale dovrebbero tendere nelle loro azioni gli uomini virtuosi secondo l'etica nicomachea di Aristotele, e che viene descritto da Galeno nel *De temperamentis*, come lo stato d'animo ugualmente distante dai due estremi: σύμμετρον ὅπερ ἑκατέρου τῶν ἄκρων ἀπέχει» (trad. it. p. 10).

Nel trattato di Aristotele *Del cielo* si attesta un'antica considerazione di filosofia della natura per cui l'idea di simmetria è utilizzata in considerazioni cosmologiche vicine per spirito ai modelli moderni dell'Universo omogeneo. Scrive Aristotele: «Se nulla fa sí che questo [il moto della Terra] sia piuttosto verso il basso che verso l'alto, e se l'aria che sta sopra non impedisce il moto verso l'alto, neppure quella che si trova sotto la terra impedirà il moto verso il basso. È necessario infatti che le medesime cause provochino nelle stesse cose eguali effetti» [295a, 26-29].

Questa idea di simmetria, che nella traduzione russa diventa подобие 'similarità' (nell'originale greco διὰ τὴν ὁμοιότητα), è riferita ad Anassimandro secondo la seguente rappresentazione del mondo. La Terra si trova al centro d'un involucro sferico di fuoco, senza appoggiarsi su alcunché: non c'è bisogno di appoggio, giacché a causa della simmetria sferica la Terra non ha alcun motivo di muoversi in alcuna direzione. Per noi è difficile valutare tutta l'originalità che questa concezione aveva quando fu elaborata per la prima volta: la Terra non era sostenuta da una tartaruga, ma dalla logica. Fra l'altro la Terra stessa, per

Anassimandro, non aveva forma sferica, ma di cilindro circolare la cui altezza era tre volte minore del diametro della base. Però in questo modo la direzione della verticale rompe chiaramente la simmetria totale dell'Universo.

L'idea che «la conseguenza non può essere meno simmetrica della causa» sorse nel XX secolo sotto il nome di principio di Curie. Probabilmente è un caso particolare di qualche archetipo di pensiero teoretico: in Anassimandro, ad esempio, è ancora chiara la sua connessione genetica con quel che, piú tardi, sarà detto «principio di ragione sufficiente» (si veda l'allegoria dell'asino di Buridano). Negli ultimi anni, nella fisica teorica, ha acquisito un significato particolare la concezione di «rottura spontanea della simmetria». Situazioni in cui «la conseguenza è meno simmetrica della causa» si verificano in un'ampia classe di modelli teorici. Il meccanismo di rottura della simmetria può essere un'instabilità particolare dello stato simmetrico in relazione a qualsivoglia piccola perturbazione esterna. Ciò sarà esaminato in dettaglio in seguito.

La connessione fra il contenuto del termine 'simmetria' e le idee di ugual peso e commensurabilità, nella concezione moderna, si è indebolita. La matematizzazione dell'idea di simmetria ha portato in primo piano la rappresentazione di una ripetibilità regolare e quindi d'un gruppo di trasformazioni in sé d'un oggetto simmetrico. A questo aspetto della simmetria si riferiscono le idee dell'antichità relative alla classificazione dei poliedri regolari e le loro connessioni con gli «elementi» di cui è costituito l'universo. La classificazione dei poliedri regolari è esposta da Euclide; il suo autore si pensa essere stato Teeteto. Quattro elementi, fuoco, aria (etere), acqua e terra, dice Empedocle. Aristotele aggiunse il quinto. Platone avanzò l'idea per cui le differenze fra gli elementi (o στοιχεῖα 'lettere dell'alfabeto' [*Teeteto*, 201d-202e]) si spiegano con la forma delle particelle da cui sono costituiti questi elementi: fuoco-tetraedro, aria-ottaedro, acqua-icosaedro, terra-cubo. Per quanto riguarda il quinto poliedro, il dodecaedro, «dio se ne giovò per decorare l'universo» [*Timeo*, 55c].

Senza dubbio la teoria di Platone appartiene a un diverso sistema di pensiero rispetto alle concezioni della scienza naturale contemporanea. Ciò nonostante nella struttura dei concetti di teoria quantistica dei campi si trovano frammenti analoghi alle concezioni di Platone. La teoria della simmetria SU(3) (e i suoi miglioramenti), che ha condotto alla concezione dei quark, confronta famiglie di particelle elementari osservabili sperimentalmente (le «lettere» dell'universo) con forme geometriche simmetriche di dimensione finita che sono dette «diagrammi di peso». Un diagramma di peso è la collezione di un numero finito di vettori (o di punti) che non appartengono allo spazio tridimensionale fisico, ma a uno spazio astratto di operatori che agiscono nello spazio dei gradi di libertà interni del sistema quantistico. Questi gradi di libertà (come l'isospin) si dicono interni perché non sono legati alla cinematica spazio-temporale, a differenza delle coordinate, dell'impulso o del momento angolare. Diagrammi di peso corrispondenti a certe n-ple sono rappresentati, ad esempio, nell'articolo «Particella» in questa stessa *Enciclopedia*, X, pp. 464-65 (cfr. anche § 2).

In tempi recenti la storia matematica della simmetria è legata al XIX secolo; il tema della simmetria riceve cosí voce autonoma in ritardo; l'analisi dell'infi-

nitamente piccolo occupava tutti. Evariste Galois, morto nel 1832, fondò la teoria della simmetria delle equazioni algebriche in una indeterminata. Da prototipo utilizzò il lavoro di Gauss sulla teoria dei campi ciclici, cioè, se si vuole, dei poligoni regolari piani. Un poligono regolare di n lati possiede una simmetria evidente: si trasforma in sé per rotazioni di angoli multipli di $2\pi/n$ attorno al centro. Si può anche notare la presenza di n assi di simmetria. La scoperta eccezionale di Gauss consisteva nel fatto che per la costruzione di un poligono regolare con riga e compasso non era tanto importante questa evidente simmetria, ma un'altra, nascosta. Si rappresentino i vertici del poligono con le radici n-esime dell'unità $e^{2\pi i(k/n)}$ sul piano complesso. Si considerino le sostituzioni di queste radici (cioè dei vertici) del tipo $\sigma_m : e^{2\pi i(k/n)} \mapsto e^{2\pi i(km/n)}$, dove il numero $0 \leq m < n$ è intero e primo con n. Risulta $\varphi(n) = n \prod_{p|n} (1-p^{-1})$ (p primo) il numero di sostituzioni, costituenti un gruppo che, nel caso generale, non è contenuto nel gruppo delle riflessioni e rotazioni. Questo gruppo non conserva la distanza fra vertici, ma la struttura delle operazioni algebriche di somma e di prodotto, sull'insieme di tutti i numeri complessi che si possono ottenere dalle radici dell'unità applicando queste operazioni. La forma della «simmetria nascosta» come della «simmetria spontaneamente distrutta» divenne una metafora matematica popolare della moderna fisica teorica.

I lavori di Galois, per motivi storici casuali, rimasero in ombra per molti anni. Il successivo sviluppo della teoria della simmetria come teoria dei gruppi e delle loro azioni è legato nella concezione dei posteri ai nomi di Klein, Jordan e Lie in primo luogo. Il *Traité des substitutions et des équations algébriques* di Jordan, che ebbe tanta influenza, apparve nel 1870. L'atmosfera in cui lavoravano e comunicavano questi tre matematici non soltanto permise la scoperta di una grande quantità di fatti matematici nuovi e la sistematizzazione di vecchi, ma anche la formazione di un'ideologia. La piú nota formulazione di questa ideologia fu il programma di Erlangen pubblicato da Klein [1872].

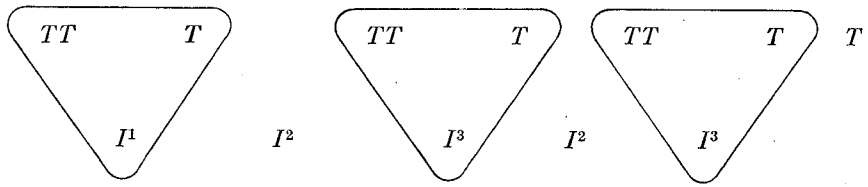
Brevemente, Klein inizia a definire il gruppo delle simmetrie di diverse teorie geometriche (euclidea, proiettiva, della geometria iperbolica). All'inizio questo gruppo serve da classificatore della teoria, ma in seguito Klein cambia punto di vista e propone di considerare che il gruppo, dato insieme alla sua azione sullo spazio, definisca in maniera essenziale tutto il contenuto della geometria corrispondente in quanto teoria delle proprietà (delle figure) invarianti per le azioni di questo gruppo.

Nei lavori di Lie (il corso di Lie-Engel è del 1888-93) gli stessi gruppi di trasformazioni di carattere geometrico, compresi quelli a dimensione infinita, diventano l'oggetto di studi coi metodi dell'analisi. Parallelamente si sviluppa la tecnica del «calcolo assoluto», vale a dire la descrizione dell'analisi locale sulle varietà in maniera indipendente dalla scelta del sistema di riferimento. Questa tecnica sarà chiamata a svolgere un ruolo importante nella teoria generale della relatività di Einstein (1916). L'idea di cambiamento delle coordinate può essere collegata all'idea di trasformazione di simmetria realizzabile fisicamente, ad esempio di *boost*, che accelera l'osservatore e l'obbliga a condurre le misura-

zioni in un sistema inerziale variabile. Allora si pone appieno il problema relativo a ciò che non dipende da questo cambiamento («invarianza»), o almeno ne dipende in maniera semplice («covarianza»). Questi invarianti e covarianti sono essenzialmente leggi fisiche.

Ma l'idea di cambiamento di coordinate può anche essere collegata globalmente alla sfera concettuale. Gli arbitrari cambiamenti di coordinate, differenziabili e con inversa differenziabile in una regione locale dello spazio-tempo, che Einstein considerava nella formulazione del suo «principio generale di relatività» non corrispondono ad alcuna trasformazione di simmetria fisicamente realizzabile. Il «principio generale di relatività» è una richiesta di simmetria applicata al linguaggio in cui si formula la legge fisica, molto maggiore della legge stessa. Altro affare è che le rappresentazioni intuitive legate a questo principio abbiano potuto svolgere, nella storia della scienza, un ruolo poco dipendente dal contenuto formale.

Nella coscienza creatrice l'idea di simmetria si riferisce a una regione limitata dall'attività sistemica, razionalizzatrice, e alla percezione estetica. Quest'idea ha la capacità di comprendere attivamente i meccanismi emozionali della coscienza. L'arte, che crea strutture spazio-temporali ritmiche, e la scienza, che le coglie in un materiale empirico non ordinato, s'incontrano qui su un territorio comune. Ecco come uno studioso moderno descrive l'*Invenzione a tre voci* in fa minore di Bach: «Nella struttura di questa *Invenzione*, scritta in contrappunto triplice e distinta da una ricchissima cromaticità, entrano dieci volte un tema emergente e tre intermedi, di cui uno compare solo una volta e gli altri due, ciascuno due volte... Le cinque comparse dei tre intermedi (1+2+2) e le dieci comparse del tema formano congiunte una simmetria molto rigorosa... Ogni intermedio si trova fra due comparse del tema delle quali una è doppia, e l'altra semplice, unica... dopo la comparsa di *tutti e tre* gli intermedi compare il secondo e poi il terzo... Con le proprie due comparse il secondo intermedio separa l'*Invenzione* in tre segmenti omogenei



i cui ultimi due sono assolutamente identici, mentre il primo si distingue dai due restanti soltanto per un intermedio» [Gerškovič 1979, pp. 44-45].

In questo testo la simmetria si rivela come principio formatore. Prendendo la teoria matematica della simmetria in una concezione ampia, si potrebbero introdurre in essa anche quelle concezioni in cui è importante l'idea di generazione ricorsiva di una struttura complessa per mezzo di un piccolo numero di regole. Il recente libro di Hofstadter e altri [1980] è una raccolta eccellente di variazioni diverse su questo tema.

1.2. Simmetria e leggi fisiche.

Le forme primitive della simmetria sono suggerite da osservazioni nelle scienze naturali.

Gli oggetti della biologia permettono di distinguere alcuni aspetti di simmetria spaziale di rotazione, che di solito sono funzionali alla crescita dell'organizzazione. Una sfera solida è dotata di simmetria rispetto a qualunque rotazione attorno al centro, vale a dire del gruppo SO(3) delle rotazioni nello spazio euclideo tridimensionale. Un cerchio solido è simmetrico rispetto alle rotazioni piane. Questi due aspetti della simmetria si realizzano approssimativamente per alcuni protozoi in sospensione nell'acqua e rispettivamente per protozoi che conducono una forma di vita sul fondo.

Organismi come le stelle marine corrispondono a un gruppo finito di rotazioni piane. Ma piú diffusa è la simmetria speculare approssimata degli organi esterni, rispetto alla riflessione destra/sinistra.

È naturale considerare le proprietà di simmetria della forma esterna in connessione con le caratteristiche delle sue interazioni col mondo esterno: col tipo di ambiente in cui abita, col carattere della localizzazione, col tipo di alimentazione. Un organismo a simmetria destra/sinistra – con eccezione della testa sulla quale sono impressi la bocca e gli occhi – è un organismo-vettore, la cui struttura è idonea ad essere attiva.

La simmetria degli organi interni è determinata parzialmente dalla simmetria esterna, ma ammette notevoli eccezioni. Negli ultimi anni si è presa molto in esame l'asimmetria funzionale, specifica dell'uomo, delle due metà del cervello, la destra e la sinistra, la quale è correlata con l'asimmetria funzionale delle mani destra e sinistra, ma è relativa a un livello molto piú profondo di attività del sistema nervoso centrale. Secondo Pribram, l'emisfero sinistro elabora l'informazione, per molti aspetti in maniera analoga a una calcolatrice numerica, mentre l'emisfero destro funziona piú rapidamente secondo i principî dei sistemi ottici e olografici di elaborazione dei dati.

Nella struttura non viva dei cristalli si trovano, probabilmente, i primi oggetti la cui ricerca richiese una teoria matematica. Un risultato di questa teoria fu la formazione di un elenco completo di 230 gruppi cristallografici. Questo elenco si ottiene come risultato della soluzione dei seguenti problemi:

a) Descrizione di tutti i gruppi finiti di rotazioni (inclusa la riflessione) dello spazio tridimensionale, vale a dire dei sottogruppi finiti di O(3) a meno di coniugio in O(3). Questi gruppi sono numerosi: vi sono due serie infinite, essenzialmente perché sul piano esistono poligoni regolari con un numero arbitrario di lati, e le rotazioni attorno a un asse passante per il loro centro possono essere d'ordine qualsiasi.

b) La descrizione di quei gruppi del precedente elenco che trasformano in sé un reticolo di rango 3, vale a dire un sistema discreto di vettori isomorfo a $\mathbf{Z}^3$ come gruppo di traslazioni. Questi gruppi sono 32. Si può qui osservare che i gruppi di simmetria dei poliedri regolari non permettono di distinguerne alcuni: il cubo e l'ottaedro, come anche il dodecaedro e l'icosaedro, hanno le stesse simmetrie. La presenza di un reticolo invariante rende naturale anche una diversa relazione di equivalenza sull'insieme dei gruppi finiti di rotazioni: coniugio a meno di trasformazioni invertibili intere. Questa classificazione risulta piú sottile e conduce a 73 classi d'equivalenza.

c) Infine, le nominate 230 classi sono classi di gruppi contenenti sia le rotazioni, sia le traslazioni; il gruppo delle traslazioni consiste precisamente negli spostamenti di vettori interi e la classificazione si effettua a meno di coniugio per mezzo di trasformazioni invertibili intere.

La simmetria dei cristalli fu scoperta durante le ricerche macroscopiche delle loro proprietà ottiche, meccaniche e chimiche. La struttura matematica dei gruppi cristallografici fu un argomento potente, ma non il principale, per capire che queste proprietà riflettono la struttura di reticoli microscopici, nei nodi dei quali si trovano atomi di cristalli componenti la materia. Dopo l'elaborazione dei metodi di ricerca roentgenografica questa comprensione ricevette una base sperimentale.

Si passerà ora a una breve rassegna del ruolo che le concezioni di simmetria svolgono nelle rappresentazioni contemporanee della fisica. La fisica ha in molto rotto con le forme di «simmetria visibile» di cui si è detto finora. Le rappresentazioni simmetriche si collegano alle idee di riproducibilità, indipendenza dall'osservatore, leggi di conservazione, principî di media statistica e molti altri.

Secondo l'osservazione del fisico americano Lee [1970] l'idea di simmetria si può collegare all'idea di «inosservabilità di alcune grandezze fondamentali». Per chiarire questo pensiero, Lee porta il seguente esempio.

Si consideri l'energia di interazione V di due particelle come funzione delle loro posizioni $\boldsymbol{x}, \boldsymbol{y}$. Il postulato fisico per cui la posizione assoluta non è osservabile conduce all'affermazione che V deve essere funzione soltanto della differenza $x-y$, in modo che ogni spostamento spaziale $x \mapsto x+a$, $y \mapsto y+a$ non cambi questa differenza, e viceversa la differenza è l'unico invariante dello spostamento. Allo stesso tempo si ha la legge di conservazione del momento totale di quantità di moto, la cui velocità di variazione col tempo è $-(\nabla_x+\nabla_y)V$.

Da un punto di vista rigorosamente logico tutta questa considerazione è vulnerabile. In effetti presuppone a priori la possibilità di costruire la «differenza» $x-y$, vale a dire postula l'esistenza e la struttura del gruppo di spostamenti spaziali. Né l'una né l'altra segue in alcun modo dal postulato sull'inosservabilità delle posizioni assolute a meno che questo postulato non venga precisato per mezzo della scelta di una struttura matematica, in questo caso d'un modello galileiano di spazio-tempo. Dal punto di vista matematico, ad esempio, anche i punti di una sfera non hanno «posizioni assolute», ma la loro differenza sulla sfera non è definita e una funzione della coppia di punti, invariante rispetto alle rotazioni della sfera, deve essere descritta in tutt'altro modo. Pertanto

l'esempio di Lee va considerato come esplicazione dell'idea di «inosservabilità delle grandezze fondamentali», piuttosto che della simmetria. In effetti, all'interno dell'apparato matematico della fisica la rappresentazione dell'inosservabilità è espressa in maniera incomparabilmente meno definita dell'idea di simmetria, che ha un formalismo sviluppato e numerose applicazioni effettive. D'altro lato, nelle discussioni semintuitive, filosofiche, assiologiche, la «inosservabilità» si presenta come una concezione molto preziosa. Nella tradizione filosofica europea l'idea di inosservabilità si può ricondurre al mito della caverna di Platone. Negli esperimenti ideali di Galileo l'inosservabilità di un movimento inerziale assoluto (a differenza di uno relativo) viene utilizzata ormai completamente nello spirito moderno. Le concezioni positiviste, popolari nel XIX secolo, fecero dell'«inosservabilità» uno spauracchio; i tentativi di espellere dalle teorie le quantità inosservabili occuparono studiosi molto seri, e questi tentativi giunsero talvolta a importanti risultati. Tuttavia le rappresentazioni del posto preciso di ciò che è inosservabile nella struttura di una teoria scientifica furono spesso indefinite. Per questo l'idea di una connessione fra inosservabilità e simmetria, almeno in fisica, richiede una maggiore considerazione.

Proseguendo la propria analisi dei principî della simmetria, Lee produce il seguente elenco, presente in pressoché tutti i manuali di teoria quantistica dei campi:

1) simmetrie di traslazione, che conducono alle statistiche di Bose-Einstein e di Fermi-Dirac;
2) trasformazioni continue dello spazio-tempo: traslazioni, rotazioni, accelerazioni, ecc.;
3) trasformazioni discrete: inversione spaziale, inversione temporale; coniugazione di carica, che scambia di posto le particelle e le antiparticelle;
4) trasformazioni unitarie, che comprendono: U_1, la simmetria che conduce alle leggi di conservazione della carica, del numero barionico e dei numeri leptonici; SU(2), simmetria (isospin); SU(3), simmetria (di colore); ecc.

Proseguendo questo elenco di concezioni, che ora si valutano piú attivamente, si possono aggiungere:

5) simmetrie di gauge; e anche
6) supersimmetrie.

Si torni alla considerazione delle simmetrie spazio-temporali. La prima cosa che si può dire è che, dopo il sorgere della teoria generale della relatività, il contenuto stesso di queste simmetrie è divenuto dipendente dal modello. Il modello generale einsteiniano dello spazio-tempo, cioè la varietà quadridimensionale liscia con pseudometrica $g_{\mu\nu}dx^\mu dx^\nu$ che soddisfa alle equazioni di Einstein con tensore di energia-impulso di aspetto qualsiasi, usualmente non ha alcuna simmetria. In altri termini, non esistono trasformazioni biunivoche non banali di questa varietà in sé, le quali conservino la pseudometrica. Si ricordi che in questi modelli non esiste né tempo assoluto, né spazio assoluto in un dato momento di tempo.

Negli spazi generali di Einstein si sceglie tutto un repertorio di modelli cosmologici. La scelta di questi modelli è dettata, naturalmente, sia dai dati osservabili sia dall'accessibilità per la ricerca matematica. Entrambe le condizioni percorrono, in primo luogo, gli spazi nei quali si ha una qualche simmetria. Spesso la simmetria è introdotta in forma infinitesimale: come esistenza di campi vettoriali che conservano la metrica (nel senso che la sua derivata di Lie lungo il campo tende a zero). Tali campi vettoriali costituiscono un'algebra di Lie. Come sostituto del gruppo delle traslazioni spaziali $\mathbf{R}^3$ (assolute o connesse con un sistema di riferimento inerziale nello spazio di Minkowski) si può usare allora un'algebra di Lie tridimensionale dei campi vettoriali simili del tipo descritto. Queste algebre di Lie furono classificate già da Bianchi, e appartengono a nove tipi, I-IX. Le usuali traslazioni corrispondono a un'algebra di Lie abeliana (tipo I), le rotazioni di una sfera 2-dimensionale costituiscono il tipo IX. Gli spazi di Einstein dei tipi corrispondenti furono sottoposti a studi accurati; in particolare, essi rivelano un'evoluzione ondulatoria molto interessante rispetto a movimenti indietro nel tempo fino al momento di singolarità, che modellizzano il *big bang*.

Fra i dati dell'osservazione che motivano lo studio di modelli spazio-omogenei dello spazio-tempo si possono osservare due fatti fondamentali: *a*) la quasi omogeneità e l'isotropia della diffusione della materia, se la si guarda con una scala abbastanza grande; *b*) l'isotropia della radiazione residua a temperatura di circa 3 °K scoperta da Penzias e Wilson.

Il modello di spazio-tempo, utilizzato al di fuori delle scale cosmologiche, ad esempio in fisica atomica, che rimane è essenzialmente lo spazio-tempo di Minkowski M (o perfino la sua versione precedente la relatività). Ciò si può spiegare matematicamente col fatto che nelle situazioni considerate la curvatura dello spazio-tempo è tanto piccola che il suo spazio tangente fornisce localmente un'approssimazione molto buona. Questo spazio tangente è proprio il mondo di Minkowski. Da un punto di vista fisico, nella misura in cui la curvatura dello spazio-tempo si può interpretare come conseguenza dell'interazione gravitazionale, il fatto che sia piccola si spiega con la particolare debolezza dell'interazione gravitazionale in confronto con quelle forti e quelle elettromagnetiche che determinano essenzialmente le proprietà della materia per scale da laboratorio. Il ruolo dell'interazione gravitazionale su scala cosmologica diventa prevalente grazie al fatto di essere universale, agente a distanza (a differenza di quella forte) e non avere «cariche di diverso segno», che potrebbero compensarsi (a differenza di quella elettromagnetica).

Lo spazio-tempo di Minkowski possiede un gruppo fondamentale di simmetria: il gruppo di Poincaré, che è generato dal gruppo quadridimensionale delle traslazioni $\mathbf{R}^4$ e dal gruppo delle rotazioni che conservano la metrica di Minkowski, vale a dire il gruppo di Lorentz. Il gruppo completo di Lorentz ha quattro componenti connesse. Se nello spazio di Minkowski si fissa una base ortogonale con tre componenti spaziali e una componente temporale ($\mathbf{e}_0$, $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$), allora le trasformazioni di cambiamento di segno del tempo T ($x^0 \mapsto -x^0$, $x^a \mapsto x^a$, $a = 1, 2, 3$), di inversione spaziale P ($x^0 \mapsto x^0$, $x^a \mapsto -x^a$, $a = 1, 2, 3$) e la

loro composizione PT appartengono a componenti diverse. La terza trasformazione dell'elenco di Lee, la coniugazione di carica C, s'interpreta meno direttamente in termini di spazio-tempo. Per introdurla occorre preliminarmente osservare che la componente connessa dell'unità L_0 del gruppo di Lorentz non è semplicemente connessa: il suo ricoprimento universale è il gruppo $\mathrm{Sl}(2,\mathbf{C})$, il nucleo del ricoprimento consiste di (± 1). L'isomorfismo di $\mathrm{Sl}(2,\mathbf{C})/(\pm 1)$, cioè del gruppo delle trasformazioni razionali lineari della sfera di Riemann $\mathbf{CP}^1$, con L_0 si esplica facilmente in maniera geometrica: anche L_0 agisce sulla sfera, precisamente sulla «sfera assoluta» dell'origine delle coordinate di L_0, vale a dire sull'insieme delle direzioni uscenti dall'origine delle coordinate.

Quello spazio complesso bidimensionale su cui agisce il gruppo $\mathrm{Sl}(2,\mathbf{C})$, ricoprente L_0, si dice spazio (a due componenti) degli spinori S_0 (nel punto iniziale). Accanto ad esso occorre considerare anche lo spazio complesso-coniugato $\bar{S}_0$; esso corrisponde a una seconda possibile orientazione della sfera assoluta, e queste orientazioni si trasformano l'una nell'altra invertendo il tempo T, scambiando i poli dell'assoluto passato e dell'assoluto futuro. Associando a ogni punto M il proprio spazio spinoriale, si ottiene su M un fibrato spinoriale quadridimensionale $S \oplus \bar{S}$ sul quale si può definire in maniera naturale l'azione di $\mathrm{Sl}(2,\mathbf{C})$, P e T, compatibile con le azioni su M. L'operazione di coniugazione di carica C porta in ogni punto di M lo spinore $s \in S$ nello spinore $\bar{s} \in \bar{S}$, e $\bar{s} \in \bar{S}$ in $s \in S$. Un'importante interpretazione fisica di C è possibile soltanto entro la descrizione quantistica della materia (a differenza della descrizione non-quantistica dello spazio-tempo a cui ci si è limitati finora). Dirac introdusse il fibrato spinoriale per la necessità della descrizione quantistica relativistica di una particella carica con spin $1/2$, ad esempio un elettrone. La funzione ψ di una simile particella, nella teoria di Dirac, è una sezione del fibrato spinoriale $S \oplus \bar{S}$. Lo sviluppo di Fourier di questa sezione permette di separare due componenti: a frequenza (o energia) positiva e negativa; la loro differenza è assoluta e i nomi, naturalmente, sono connessi a una scelta dell'orientazione temporale dello spazio di Minkowski. Quando non era ancora formulata la concezione delle antiparticelle (dei «buchi» nella versione originaria), si riteneva che solo le sezioni a frequenza positiva avessero significato fisico. Proprio per questo la comparsa di componenti a frequenza negativa nelle soluzioni dell'equazione di Dirac, che descrivevano l'elettrone nel campo elettromagnetico esterno (classico), era considerata un inspiegabile difetto della teoria.

Dopo la scoperta del positrone (antiparticella dell'elettrone) risultò chiaro che l'operatore C attesta una fondamentale simmetria fra particelle e antiparticelle nella teoria semiclassica di Dirac, e le sezioni del fibrato spinoriale modellizzano uno spazio di stati di un unico campo elettrone-positrone. L'indivisibilità delle corrispondenti componenti di Fourier riceve una profonda spiegazione in una teoria maggiormente conseguente, dove anche il campo elettromagnetico è trattato come campo quantistico e il campo dei positroni si sottomette a quanti secondari, cosa che permette di tener conto dei processi in cui il numero delle particelle non si conserva.

Una conseguenza di questa teoria dell'elettrodinamica quantistica è la rappresentazione del vuoto, nel quale si verifica una nascita spontanea e l'annichilazione delle coppie virtuali elettrone-positrone. In un campo abbastanza fortemente classico (cioè multifotonico quantistico), le particelle virtuali possono trasformarsi in reali. Uno «smussamento», una riflessione semiclassica di ciò è la superposizione delle due componenti del campo spinoriale nell'originaria trattazione di Dirac.

Questa breve descrizione della storia della simmetria di C può servire da esempio della forza predittrice delle considerazioni sulla simmetria già a livello quantistico. Da un punto di vista puramente matematico si può considerare che con ciò sia giunta a compimento logico l'idea di Einstein relativa al fatto che il gruppo di Lorentz, compreso come gruppo delle simmetrie delle equazioni di Maxwell, sia il gruppo universale delle simmetrie per tutte le leggi della natura. La circostanza per cui l'interpretazione fisica di un ricoprimento semplicemente connesso del gruppo di Lorentz (o la sua «doppia rappresentazione», come dicono i fisici) $\mathrm{Sl}(2,\mathbf{C})$ richiede già la teoria quantistica e, soprattutto, l'introduzione della prima osservabile realmente quantistica – lo spin e ancor di piú la previsione dell'antimateria –, è un'illustrazione eccezionale della forza delle considerazioni della simmetria.

La storia delle applicazioni C, P, T, con ciò, è ben lungi dall'esser finita. In termini generali, sono le simmetrie dello spazio di stati di qualunque teoria quantistica del campo che opera con uno spazio-tempo piano e tiene conto delle interazioni elettromagnetiche e degli spin. Ma la dinamica può non conservare le simmetrie dello spazio delle fasi per diversi motivi: *a*) le leggi stesse del moto possono non essere invarianti rispetto alle simmetrie; *b*) può avvenire che le leggi del moto siano invarianti, ma che i fenomeni osservati non lo siano a causa di qualche condizione iniziale o al contorno (non controllate dalla teoria); *c*) lo stesso a causa di fenomeni d'instabilità (queste alterazioni della simmetria si è preso a chiamarle spontanee).

L'alterazione dinamica della simmetria posseduta dallo spazio delle fasi è soggetta a una dimostrazione sperimentale. La scelta fra le alternative *a*), *b*), *c*) (o una loro opportuna combinazione) per spiegare questa asimmetria è diventata un conseguente problema della teoria.

Un esperimento mostra che in quei processi in cui le interazioni elettromagnetica e forte possono essere considerate indipendenti dalle altre, le simmetrie C e P non sono alterate. Ciò significa che per ogni processo del genere può essere realizzato un processo «specularmente simmetrico», nel quale le particelle sono sostituite con antiparticelle. La «realizzazione» consiste nel preparare le condizioni iniziali corrispondenti e nel misurare le probabilità degli stati finiti. Solo queste probabilità sono P- (oppure C-) invarianti.

Nel 1957 si scoprí che nei processi in cui intervengono interazioni deboli le simmetrie spaziali e di carica si alterano. Il K-mesone neutro a lunga vita K_L^0 è uno stato C- e P-invariante. I suoi due modi possibili di decomposizione $K_L^0 \to e^+\pi^-\nu_e$ e $K_L^0 \to e^-\pi^+\nu_e$ si trasformano l'uno nell'altro per mezzo dell'operatore C. Il rapporto fra le loro probabilità è uguale a 1,003. Nella decompo-

sizione dei mesoni K^0 si alterano anche la P-invarianza, la CP-invarianza combinata e la T-invarianza.

Le concezioni moderne all'interno delle quali si costruisce la teoria quantistica del campo che tiene conto dell'asimmetria delle interazioni deboli, si basano sull'idea di simmetria di gauge. La sua comparsa si può ancora far risalire all'esempio dell'originaria teoria di Dirac. Si ricordi che la costruzione degli spinori era connessa all'introduzione dello spazio delle rappresentazioni del gruppo $Sl(2,\mathbf{C})$, ricoprente il gruppo di Lorentz. È lo spazio dei gradi spinoriali di libertà. Nel formalismo della meccanica quantistica si assume che lo spazio degli stati di un sistema quantistico isolato sia lo spazio dei raggi complessi (sottospazi unidimensionali) di uno spazio complesso di Hilbert, vale a dire, in linguaggio matematico, lo spazio proiettivo associato. Se si rappresenta un raggio contenente il vettore ψ di modulo 1, allora ψ si può sostituire con $e^{i\vartheta}\psi$, senza cambiare stato. «Il fattore non osservabile di fase» $e^{i\vartheta}$ non occorrerebbe considerarlo, se a ciò non obbligassero i principî di descrizione dell'unione dei sistemi quantistici: lo spazio degli stati di un'unione di sistemi è connesso al prodotto tensoriale degli spazi di stati originari – operazione che non si può formulare soltanto in termini di raggi, finché l'interazione dei sistemi (e anche la statistica quantistica) conduce a superposizione. A sua volta ciò rende osservabili effetti di «differenza di fase».

La simmetria di gauge dell'equazione di Dirac si manifesta nel fatto che questa equazione è invariante rispetto a un simultaneo cambiamento del campo $\psi(x)$ col campo $e^{i\vartheta(x)}\psi(x)$ («fattori di fase indipendenti in punti diversi») e del potenziale elettromagnetico A_μ con $A_\mu - \partial_\mu\vartheta$, $\partial_\mu = \partial/\partial x^\mu$ (ciò non cambia le componenti del tensore $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, che corrispondono a un'unica caratteristica essenziale, locale, del campo).

Tuttavia l'invarianza di gauge può essere interpretata piú ampiamente: come principio che postula la necessità d'introdurre i campi A_μ, in grado di realizzare cambiamenti negli spazi dei gradi interni di libertà da punto a punto dello spazio-tempo. Naturalmente il campo elettromagnetico non fu introdotto in questo modo. Tuttavia il campo portante l'interazione forte (i cui quanti si chiamano gluoni) fu introdotto proprio in questo modo. Inizialmente si osservò che le forze che uniscono nel nucleo un neutrone e un protone sono quasi le stesse e si postulò che, dal punto di vista delle forze nucleari, con «esclusione» dell'interazione elettromagnetica, il neutrone e il protone si possono considerare come stati degeneri della stessa particella. Dalla loro superposizione è possibile costituire uno spazio d'isospin (in ogni punto dello spazio-tempo) bidimensionale. Il campo A_μ a valori nelle matrici (2×2), che può realizzare cambiamenti nello spazio dell'isospin, è introdotto in modo naturale come oggetto portante un'interazione. La sua lagrangiana si costruisce secondo il modello della lagrangiana dell'equazione di Maxwell, con una differenza essenziale: poiché i potenziali matriciali di A_μ, a differenza di quelli di Maxwell, non sono commutativi, per la conservazione dell'invarianza di gauge (rispetto alla sostituzione $\psi \mapsto e^{i\vartheta}\psi$) occorre aggiungere dei termini non lineari all'equazione di moto, come fu rivelato da Yang e Mills. In termini tecnici, i campi di gauge

non abeliani posseggono un'autoazione che rende molto piú complessa la loro teoria, e i fenomeni osservabili in tutto distinguibili da quelli elettromagnetici.

A questa descrizione schematica dei campi di gauge occorre aggiungere alcuni dettagli e precisazioni. Nei modelli moderni di interazioni forti, i neutroni e i protoni non si considerano come costituenti elementari, il loro ruolo è assunto dai quark. Corrispondentemente quei gradi interni di libertà nei cui spazi agiscono gruppi di simmetrie, non sono piú d'isospin. Negli schemi piú semplici e fondamentali svolgono il ruolo principale due gruppi: $SU(3)_f$ e $SU(3)_c$. Il primo mescola quark di tre tipi: è una simmetria approssimata, perché le masse di questi quark sono diverse. Il secondo mescola tre colori dei quark: questo gruppo si considera ora esatto, e il lagrangiano del campo dei gluoni è invariante rispetto ad esso. Si pone intensivamente lo studio del modo con cui le proprietà d'una teoria di gauge non abeliana permettono di chiarire l'assenza sperimentale dei quark in uno stato libero (confinamento del colore). L'introduzione del gruppo $SU(3)_f$ precedette, propriamente, la postulazione dei quark.

Le idee di simmetria di gauge svolsero un ruolo decisivo anche nella formazione della prima teoria unitaria delle interazioni elettromagnetiche e deboli, la cosiddetta teoria di Salam-Weinberg. Nel loro modello il fotone e tre nuove particelle $W^\pm$, Z^0, portatori di interazioni deboli, sono i quanti di un campo di gauge. I bosoni intermedi di $W^\pm$, Z^0 non sono ancora stati scoperti sperimentalmente, ma le loro proprietà sono previste abbastanza precisamente ed esperimenti per la loro scoperta sono progettati per i prossimi anni.

Si dirà infine qualche parola sulle statistiche di Fermi e di Bose-Einstein. Le particelle, le cui funzioni ψ appartengono a una potenza tensoriale pari del fibrato spinoriale, vengono dette bosoni; a una dispari, fermioni. Nell'ultimo decennio si è cominciato a considerare le trasformazioni di supersimmetria, che mescolano fermioni e bosoni. Dal punto di vista matematico ciò conduce a considerare varietà sulle quali, oltre alle consuete funzioni, sono assegnate anche funzioni di coordinate «dispari», antisimmetriche e i gruppi con queste coordinate sono supergruppi di Lie. La supersimmetria di gauge si considera oggigiorno come principio fondamentale per la costruzione di una teoria unificata di tutte le interazioni.

2. *Gruppi e loro azioni.*

2.1. Esempi e illustrazioni.

La maggior parte delle definizioni utilizzate piú avanti è stata data negli articoli «Applicazioni», «Dualità», «Strutture matematiche», «Razionale/algebrico/trascendente» in questa stessa *Enciclopedia*. Qui non le si ripeterà, ma semplicemente le si illustrerà. Sia X un insieme di n elementi, ad esempio $\{1, 2, \ldots, n\}$. S_n sia l'insieme delle applicazioni biunivoche di X in sé, o trasformazioni di X. S_n è chiuso rispetto alla composizione, la quale è associativa; in S_n è contenuta l'applicazione identica e accanto a ogni applicazione esiste l'in-

versa. In tal modo S_n è un gruppo. Analogamente l'insieme degli isomorfismi in sé (automorfismi) di un oggetto arbitrario in una qualsiasi categoria è un gruppo (cfr. il citato articolo «Applicazioni»). Il numero d'elementi, o ordine di S_n, è $n!$

Il gruppo S_n è assegnato insieme alle proprie azioni: $S_n \times X \to X$, $(g,x) \mapsto g(x)$. Ogni trasformazione $g \in S_n$ genera un sottogruppo ciclico $\{1, g, g^2, ..., g^{k-1}\}$, dove k è l'ordine di g. È comodo rappresentare un'azione g su $\{1, 2, ..., n\}$ sotto forma di collezione di cicli di tipo (125) (4) (36); una simile scrittura indica che $g(1)=2$, $g(2)=5$, $g(5)=1$, $g(4)=4$, $g(3)=6$ e $g(6)=3$. I sottoinsiemi di $\{1, 2, ..., n\}$ compresi fra parentesi sono le orbite del sottogruppo generato da g: sottoinsiemi minimali di X che il sottogruppo trasforma in sé. L'intero gruppo S_n agisce transitivamente su X, vale a dire X è costituito da un'orbita. Il sottogruppo di S_n che lascia fisso un punto $x \in X$, cioè il sottogruppo stazionario di questo punto, è isomorfo a S_{n-1}: il gruppo delle trasformazioni di $X \setminus \{x\}$. Se si cambia il punto x, allora il sottogruppo stazionario $S(x)$ si trasforma nel sottogruppo coniugato $S(y)$: $S(y) = gS(x)g^{-1}$, dove $g \in S_n$ è un elemento qualsiasi per il quale $g(y)=x$. Un sottogruppo per il quale coincidono tutti i suoi coniugati si dice normale. I sottogruppi normali del gruppo G sono esattamente i nuclei degli omomorfismi di G in un altro gruppo H: $G \to H$. Un gruppo privo di sottogruppi normali non banali (diversi da $\{1\}$ e da $\{G\}$) si dice semplice. Per $n \geq 5$ il gruppo S_n ammette un unico sottogruppo normale non banale A_n, costituito dalle trasformazioni pari, le quali lasciano invariato il polinomio $\prod_{1 \leq i < j \leq n} (x_i - x_j)$ (qui $X = \{x_1, x_2, ..., x_n\}$). Il sottogruppo A_n è già semplice.

Di nuovo, supponendo $X = \{x_1, x_2, ..., x_n\}$ è possibile definire l'azione di S_n sullo spazio lineare dei polinomi omogenei F di grado $d \geq 1$ in $x_1, ..., x_n$, a coefficienti complessi: $T_g F(x_1, ..., x_n) = F(g^{-1}(x_1), ..., g^{-1}(x_n))$. Ogni operatore T_g è un'applicazione lineare, e $T_{gh} = T_g T_h$ per ogni $g, h \in S_n$. Una tale azione lineare su uno spazio lineare si dice rappresentazione lineare del gruppo.

Il gruppo $G = \mathrm{Gl}(L)$ di tutte le applicazioni lineari invertibili $g: L \to L$, dove L è uno spazio lineare a dimensione finita su $\mathbf{R}$ oppure su $\mathbf{C}$, è uno spazio topologico ed è anche una varietà differenziabile. Le applicazioni $G \times G \to G$, $G \times L \to L$ che definiscono la legge di composizione e l'azione, sono morfismi delle corrispondenti categorie. In tal modo i gruppi lineari $\mathrm{Gl}(L)$ sono il principale esempio di oggetti gruppo generali in una categoria e di loro azioni categoriali: i gruppi di Lie.

Si consideri la categoria delle supervarietà differenziabili. I suoi oggetti sono coppie formate da una varietà differenziabile e da un fascio di anelli che è localmente isomorfo all'algebra esterna di un fibrato di dimensione finita. I morfismi sono gli abituali morfismi di spazi anellati. Gli oggetti gruppo in questa categoria si dicono supergruppi di Lie.

Il gruppo di Lie G agisca sulla varietà X. Quest'azione può essere un modello d'evoluzione temporale nello spazio delle fasi del sistema, o di azione di un gruppo di simmetria. I seguenti problemi relativi alla struttura dell'azione sono fondamentali:

a) La struttura dell'insieme delle orbite, la topologia delle loro mutue posizioni, le orbite di misura massimale e minimale, che attraggono o respingono gli insiemi.

b) Il legame dell'azione con altre strutture, ad esempio una misura su X. L'esistenza di misure invarianti, la struttura degli insiemi instabili invarianti.

c) La decomposizione delle rappresentazioni di G in spazi funzionali su X di diverso tipo (analitici, lisci, misurabili, ecc.).

Nello studio di questi problemi svolge un ruolo importante la teoria spettrale degli operatori: la piú importante caratteristica dell'azione di un operatore su uno spazio lineare.

2.2. Spettro.

Caso a dimensione finita. Sia L uno spazio lineare a dimensione finita sul campo dei numeri complessi $\mathbf{C}$; $A: L \to L$ sia un operatore lineare. Si dice spettro di A il sottoinsieme $\sigma(A) \subset \mathbf{C}$ che ammette le seguenti descrizioni equivalenti:

a) $\lambda \in \sigma(A) \Leftrightarrow$ l'operatore $A - \lambda \cdot \mathrm{id}$ non è invertibile ($\mathrm{id}: L \to L$ è l'operatore identico).

b) $\lambda \in \sigma(A) \Leftrightarrow$ esiste un vettore non nullo $\mathbf{l} \in L$ tale che $A(\mathbf{l}) = \lambda \mathbf{l}$. Se l'ultima condizione è verificata, il numero λ si dice autovalore di A, il vettore $\mathbf{l}$ è un autovettore associato e lo spazio $\mathrm{Ker}(A - \lambda \cdot \mathrm{id})$ l'autospazio associato.

Nel caso generale degli spazi lineari qualsiasi gli autovalori appartengono allo spettro, ma non lo esauriscono necessariamente. Il fatto che ogni elemento dello spettro sia un autovalore, nel caso a dimensione finita, deriva dall'esistenza del determinante: $\det(A - \lambda \cdot \mathrm{id})$ come polinomio in λ ammette tante radici complesse (tenendo conto delle molteplicità) quanta è la dimensione di L e per ogni radice λ_i il nucleo di $A - \lambda_i \cdot \mathrm{id}$ non è banale.

Lo spettro di A è invariante rispetto al coniugio con un arbitrario operatore invertibile $B: A \mapsto BAB^{-1}$. (Questo coniugio si può considerare come un automorfismo del diagramma $A: L \to L$). Nella classe degli operatori diagonalizzabili, cioè degli operatori per cui L si decompone come somma diretta di autospazi, la coincidenza degli spettri implica il coniugio.

Le seguenti importanti classi di operatori sono diagonalizzabili:

1) Operatori con spettro semplice, cioè aventi tanti autovalori mutuamente disgiunti quanta è la dimensione di L. Poiché è possibile con una deformazione piccola quanto si voglia trasformare un arbitrario operatore in un operatore a spettro semplice, ne segue che gli operatori diagonalizzabili costituiscono un aperto ovunque denso nello spazio di tutti gli operatori.

2) Operatori che conservano il prodotto scalare non degenere, simmetrico o hermitiano. (Questi operatori si dicono rispettivamente ortogonali e unitari). Questi operatori costituiscono un gruppo di Lie (quando è fissato il corrispondente prodotto scalare).
3) Operatori autoaggiunti, cioè operatori con la proprietà $(A\mathbf{l}_1, \mathbf{l}_2) = (\mathbf{l}_1, A\mathbf{l}_2)$, dove $(\mathbf{l}_1, \mathbf{l}_2)$ è il prodotto scalare in L di uno dei due tipi considerati nel punto precedente. Questi operatori costituiscono un'algebra di Lie corrispondente a un gruppo di Lie.

La distribuzione dello spettro d'un operatore svolge un ruolo importante in numerosi problemi. Piú diffuse sono le applicazioni a equazioni di evoluzione del tipo $d\mathbf{l}/dt = A\mathbf{l}$, dove $\mathbf{l}$ è un vettore di L dipendente dal tempo t. Nel caso $\dim L = 1$, l'operatore A è semplicemente il prodotto per un numero complesso $\lambda = a + bi$ e il comportamento qualitativo di ogni soluzione dell'equazione $\mathbf{l}(t) = e^{(a+bi)t}\mathbf{l}(0)$ è definito dalla distribuzione di $a+bi$ come segue: il segno corrisponde all'evoluzione della grandezza $\mathbf{l}$: $|\mathbf{l}(t)|$ tende a zero per $t \to \infty$ se $a < 0$, e tende ad ∞ per $t \to \infty$ se $a > 0$. La parte immaginaria b fornisce la fase, o la componente ondulatoria del processo.

Il caso $a < 0$ può corrispondere a processi quali la disintegrazione radioattiva, il caso $a > 0$ a un'esplosione esponenziale. Poiché le equazioni di evoluzione lineare sono spesso soltanto descrizioni approssimate di modelli non lineari piú precisi in un intorno della posizione di equilibrio $\mathbf{l} = \mathbf{0}$, il caso $a < 0$ corrisponde a un equilibrio stabile, $a > 0$ non stabile. Il caso intermedio $a = 0$, $b \neq 0$ descrive la possibilità di movimento ciclico in un intorno del punto di equilibrio.

Se $\dim L > 1$ e l'operatore A è diagonalizzabile, l'evoluzione è il prodotto diretto di un numero finito di processi omogenei del tipo descritto. In particolare la distribuzione di $\sigma(A)$ nel semipiano $\mathrm{Re}\,\lambda < 0$ è un indice di stabilità del baricentro $\mathbf{l} = \mathbf{0}$, e nel semipiano chiuso $\mathrm{Re}\,\lambda \leq 0$ di semistabilità. Il nuovo fenomeno che si presenta qui è la possibilità che l'operatore A cessi di essere diagonalizzabile e assuma una radice multipla λ. In questo caso, anche per $\mathrm{Re}\,\lambda = 0$, è possibile un movimento di ampiezza illimitata, crescente in modo polinomiale nel tempo, vale a dire il caso semistabile può divenire instabile quando le radici coincidono.

Essenzialmente si è considerato l'operatore e^{At} connesso all'equazione di evoluzione. Un altro problema, la teoria delle catene di Markov, conduce a un'evoluzione con tempo discreto $n \mapsto A^n$. In questo problema si considera un insieme costituito da d stati di un sistema che, in successivi momenti interi di tempo, può passare casualmente da uno stato a un altro. La probabilità di passaggio dallo stato i allo stato j sia p_{ij} (dipendente solo dalla coppia di stati). Si denoti con A la matrice i cui elementi sono p_{ij}. Sia $\mathbf{l} = (l_i)$ la colonna che rappresenta la distribuzione di probabilità sull'insieme degli stati al momento $n = 0$. Dopo il tempo n si ottiene la distribuzione di probabilità $A^n\mathbf{l}$. Gli elementi della matrice A sono non-negativi e la somma degli elementi di ciascuna colonna è uguale a 1; i corrispondenti operatori si dicono stocastici. Lo spettro di un operatore stocastico è costituito dalle radici dell'unità e anche da numeri

strettamente contenuti entro il cerchio unitario. Questi ultimi contribuiscono al comportamento di $A^n\mathbf{l}$ con termini tendenti a zero per $n \to \infty$, mentre i primi determinano il comportamento asintotico del sistema.

Funzioni di operatori. In connessione coi modelli di evoluzione ci si è già imbattuti in funzioni di operatori del tipo e^{At} e A^n. Lo studio sistematico di queste funzioni di operatori negli spazi topologici permette di estendere la teoria dello spettro al caso infinito. Per i concetti utilizzati piú avanti si veda l'articolo «Dualità» in questa stessa *Enciclopedia*, V, pp. 162 sgg.

Sia A un operatore $L \to L$, dove L è uno spazio topologico, per definizione sul campo dei numeri complessi. Si tenterà di definire $f(A)$ per una classe di funzioni f piú ampia, estendendola per mezzo di alcuni procedimenti.

a) *Chiusura rispetto alle operazioni di anello.* Se $f(A)$ e $g(A)$ sono già definite, si può definire $(\alpha f + \beta g)(A)$, dove $\alpha, \beta \in \mathbf{C}$, e $(fg)(A)$ per mezzo delle formule $\alpha f(A) + \beta g(A)$ e $f(A) \cdot g(A)$ rispettivamente. In particolare, partendo dalla funzione $f(x) = x$, si può definire $p(A)$ per ogni polinomio in A.

b) *Divisione.* Se $f(A)$ e $g(A)$ sono già definite e $f(A)$ è invertibile, si può definire $(f^{-1}g)(A) = f^{-1}(A)g(A)$. In particolare per $R_\lambda(x) = (x-\lambda)^{-1}$ l'operatore $R_\lambda(A)$ è definito se $A - \lambda \cdot \mathrm{id}$ è invertibile, cioè se λ non appartiene allo spettro di A: proprio questa definizione di spettro viene assunta al posto di quella usuale. È facile mostrare che se $f(x)$ è un polinomio le cui radici non appartengono allo spettro di A, allora $f(A)$ è invertibile.

c) *Completamento topologico.* È la prima operazione in cui si utilizza essenzialmente il passaggio al limite. Si supponga di aver definito $f(A)$ per la funzione f in qualche anello K. Si supponga che L sia uno spazio normato tale che A abbia norma $\|A\| < \infty$, e che K sia dotato di una norma rispetto alla quale l'applicazione $f \mapsto f(A)$ sia continua. Allora questa applicazione si può estendere al completamento $\hat{K}$ dell'anello K rispetto alla sua norma. Utilizzando questa astrazione si può, ad esempio, definire $f(A)$ per ogni funzione continua, se $n = \dim L < \infty$ e A è diagonalizzabile, oppure per ogni funzione di A differenziabile n volte in senso ordinario.

Sono possibili altri completamenti topologici; in seguito si riporteranno solo alcuni risultati.

Operatori autoaggiunti limitati in uno spazio di Hilbert. Sia ora A un operatore limitato nello spazio di Hilbert H. Valgono i seguenti fatti:

a) Lo spettro di A è un sottoinsieme compatto, non vuoto di $\mathbf{C}$. Lo spettro dell'operatore aggiunto A^* ha la forma

$$\sigma(A^*) = \overline{\sigma(A)} = \{\bar{\lambda} \mid \lambda \in \sigma(A)\}.$$

b) Sia $f(x)$ una funzione razionale di x, priva di singolarità in $\sigma(A)$. Allora è definito l'operatore $f(A)$ e il suo spettro ha la forma

$$\sigma(f(A)) = f(\sigma(A)) = \{f(\lambda) \mid \lambda \in \sigma(A)\}.$$

c) Il numero $\sup\{|\lambda| \mid \lambda \in \sigma(A)\} = \tau(A)$ si dice raggio spettrale di A. Per ogni operatore limitato è uguale a $\lim_{n\to\infty} \|A^n\|^{1/n}$ e per uno autoaggiunto a $\|A\|$. Lo spettro di un operatore autoaggiunto è reale e, in tal modo, è contenuto nell'intervallo $[-\|A\|, \|A\|]$. Se la funzione razionale $f(x)$ non ha poli nello spettro, allora

$$\|f(A)\| = \max_{\lambda\in\sigma(A)} |f(\lambda)|.$$

Completando rispetto alla norma corrispondente, si ottiene la definizione di $f(A)$ per ogni elemento f dell'anello $\mathcal{C}[-r, r]$ delle funzioni continue nell'intervallo $[-r, r]$, dove $r = \|A\|$.

d) Per ogni operatore autoaggiunto A esiste un unico omomorfismo continuo dell'anello $\mathcal{C}[-r, r]$ nell'algebra degli operatori limitati su H: $f \mapsto f(a)$, per cui valgono le seguenti proprietà: $1 \mapsto \mathrm{id}$; $x \mapsto A$; se $f \mapsto B$, allora $\bar{f} \mapsto B^*$; infine l'omomorfismo non aumenta la norma.

Si consideri ora un'altra estensione $B[-r, r]$ dell'algebra $\mathcal{C}[-r, r]$: essa consiste delle funzioni boreliane limitate nell'intervallo $[-r, r]$. Gli elementi di quest'algebra si ottengono dalle funzioni continue con un passaggio al limite puntuale.

e) Sotto le condizioni del punto precedente, è possibile definire l'estensione a $B[-r, r]$ dell'omomorfismo descritto, con le seguenti proprietà: se (f_n) è una successione di funzioni continue, limitate uniformemente, di $\mathcal{C}[-r, r]$, e $f(x) = \lim f_n(x)$ per ogni $x \in [-r, r]$, allora $f(A) = \lim f_n(A)$.

Si darà un esempio che, come si vedrà piú avanti, è tipico. Si consideri uno spazio di misura (X, μ) e sia $H = L_2(X, \mu)$ (si veda l'articolo «Dualità» in questa stessa *Enciclopedia*). Ogni funzione $a \in L_\infty(X, \mu)$ definisce l'operatore A di prodotto per a in H. In questo caso $f(A)$ è l'operatore di prodotto per $f(a(x))$.

Operatori autoaggiunti non limitati. Se nell'esempio del paragrafo precedente si prende per a una funzione che non appartiene a $L_\infty(X, \mu)$, l'operatore di prodotto per essa può risultare definito non su tutto $L_2(X, \mu)$. Ad esempio se $X = \mathbf{R}$, μ = misura di Lebesgue, $a = x$, allora $xf \in L_2(X, \mu)$ non per ogni $f \in L_2(X, \mu)$.

Si assoceranno questi operatori A, non ovunque definiti, ai loro grafici $\Gamma_A \subset H \oplus H$; il grafico può essere un arbitrario sottospazio lineare algebrico non contenente elementi del tipo $0 \oplus h$, $h \neq 0$. La proiezione di Γ_A sul primo addendo H si dice dominio di definizione $\mathfrak{D}_A$ dell'operatore A, che definisce l'applicazione $A : \mathfrak{D}_A \to H$ per mezzo della condizione $x \oplus Ax \in \Gamma_A$.

Si definisce una involuzione t su $H \oplus H$ con la formula $t(g \oplus h) = -h \oplus g$ e s'introduce l'operatore aggiunto ad A mediante la formula Γ_{A^*} = complemento ortogonale a $t(\Gamma_A)$. Il dominio di definizione di $\mathfrak{D}_{A^*}$ può essere caratterizzato come il sottospazio $\{h \mid$ *il funzionale* $g \mapsto (Ag, h)$ *è limitato su* $\mathfrak{D}_A\}$. L'usuale uguaglianza $(Ag, h) = (g, A^*h)$ è valida per $g \in \mathfrak{D}_A$, $h \in \mathfrak{D}_{A^*}$.

L'operatore B si dice un'estensione di A, $B \supset A$, se $\Gamma_B \supset \Gamma_A$. La principale estensione di A è la chiusura, corrispondente alla chiusura del grafico Γ_A. Tuttavia essa non è sempre definita, poiché alla chiusura di Γ_A possono appartenere dei punti del tipo $0 \oplus h$, $h \neq 0$. Un operatore con grafico chiuso si dice chiuso. A^* è sempre chiuso e $(A^*)^*$, quando è definito, coincide con la chiusura di A.

Per gli operatori considerati, non necessariamente limitati, s'introduce una generalizzazione della condizione di autoaggiunzione in alcune varianti non equivalenti.

Un operatore per cui $(Ax, y) = (x, Ay)$ per $x, y \in \mathfrak{D}_A$ ovvero $A \subset A^*$, si dice simmetrico.

Un operatore per cui $\mathfrak{D}_A = \mathfrak{D}_{A^*}$ e $A = A^*$ nel comune dominio di definizione si dice autoaggiunto. Infine un operatore la cui chiusura è autoaggiunta si dice essenzialmente autoaggiunto.

Ad esempio sia A l'operatore di prodotto per x in $L_2(\mathbf{R}, dx)$ con dominio di definizione $\mathfrak{D}_A = \mathfrak{D}(\mathbf{R})$ (funzioni a supporto compatto indefinitamente differenziabili su $\mathbf{R}$). Allora A^* è l'operatore di prodotto per x con dominio di definizione $\mathfrak{D}_{A^*} = \{f \in L_2(\mathbf{R}, dx) \mid xf \in L_2(\mathbf{R}, dx)\}$. Cosí A^* è autoaggiunto, A è simmetrico ed essenzialmente autoaggiunto.

In maniera analoga, sia $A = d/dx$, $\mathfrak{D}_A = \mathfrak{D}(\mathbf{R})$, $L = L_2(\mathbf{R}, dx)$. Allora $A^* = -d/dx$ e $\mathfrak{D}_{A^*}$ consiste di quelle funzioni la cui derivata generalizzata appartiene a $L_2(\mathbf{R}, dx)$.

Gli operatori spesso sono dati con espressioni differenziali; in questo caso non è difficile verificare la simmetria dell'operatore. Se A è simmetrico, allora essere essenzialmente autoaggiunto equivale al fatto che $\mathrm{Ker}(A^* \pm i \cdot \mathrm{id}) = \{0\}$ e anche al fatto che l'immagine di $A^* \pm i \cdot \mathrm{id}$ è densa in H. Ad esempio l'operatore $i(d/dx)$ in $L_2(\mathbf{R}, dx)$ è simmetrico. L'equazione $(A^* + i \cdot \mathrm{id}) f = 0$ ha soluzioni $f = c \cdot e^{\pm x}$ che non appartengono a $L_2(\mathbf{R}, dx)$ di modo che $\mathrm{Ker}(A^* \pm i \cdot \mathrm{id}) = \{0\}$. Di conseguenza A è essenzialmente autoaggiunto.

Teorema spettrale. Eccone la variante piú semplice. Siano $A_1, A_2, \ldots, A_n$ degli operatori autoaggiunti, limitati, a due a due permutabili nello spazio di Hilbert H. Si può allora costruire un isomorfismo di H con $L_2(X, \mu)$ e delle funzioni $a_i(x) \in L_\infty(X, \mu)$ di modo che A_i diventi il prodotto per a_i.

Utilizzando questo teorema si può definire una funzione a partire da operatori autoaggiunti; non necessariamente limitati. Piú precisamente per un tale operatore A esiste un unico omomorfismo dell'algebra $B(\mathbf{R})$ delle funzioni boreliane limitate su $\mathbf{R}$ nell'algebra $\mathrm{End}\, H$, che porta 1 in id e gode delle seguenti proprietà:

a) $\varphi[(t+i)/(t-i)] = (A + i \cdot \mathrm{id})(A - i \cdot \mathrm{id})^{-1} = B$.

b) Se $|f_n(t)| \leq C$ e $f_n(t) \to f(t)$ per ogni $t \in \mathbf{R}$, allora $\varphi(f_n)$ converge fortemente a f.

La dimostrazione consiste nella costruzione di B come prodotto per una funzione misurabile e in un opportuno passaggio al limite.

Il metodo piú comodo per formulare il teorema spettrale generale consiste nell'introdurre un sistema di concetti connessi alla misura spettrale.

Sia (X, B) uno spazio con una σ-algebra di sottoinsiemi, contenente X,

e sia H uno spazio di Hilbert. Si chiama misura spettrale una funzione $\lambda: B \to \mathrm{End}\, H$ che verifica le condizioni:

a) i valori di λ sono proiettori ortogonali;
b) $\lambda(E_1 \cap E_2) = \lambda(E_1)\lambda(E_2)$; $\lambda(E_1 \cup E_2) = \lambda(E_1) + \lambda(E_2)$, se $E_1, E_2 \in B$ sono disgiunti;
c) se esiste il limite $E = \lim_{n\to\infty} E_n$, allora $\lambda(E) = s\text{-}\lim \lambda(E_n)$.

Esempio caratteristico: sia μ una misura σ-additiva su B, $H = L_2(X, \mu)$, $\lambda(E) =$ operatore di prodotto per la funzione caratteristica di E.

Grazie al metodo di Lebesgue, è possibile definire l'integrale associato a una misura spettrale; i suoi valori sono operatori.

Il teorema generale sulla decomposizione spettrale di un operatore autoaggiunto A, non necessariamente limitato, può ora essere formulato come segue. Esiste un'unica misura spettrale λ sui sottoinsiemi di Borel B, per cui $A = \int_{-\infty}^{+\infty} x\, d\lambda(x)$. Inoltre, per ogni funzione boreliana limitata f su $\mathbf{R}$, $f(A) = \int_{-\infty}^{+\infty} f(x)\, d\lambda(x)$.

Se f è misurabile, ma non limitata, si può definire questo integrale come operatore con dominio di definizione $D_{f(A)} = \{\xi \in H \mid \int f(x)^2\, d\lambda_\xi(x) < \infty\}$, dove $\lambda_\xi(E) = (\lambda(E)\xi, \xi)$.

Un'analoga caratterizzazione degli operatori autoaggiunti di dimensione finita come generatori infinitesimali del gruppo unitario è il teorema di Stone.

Sia $\{V(t)\}$, $t \in \mathbf{R}$, una famiglia di operatori unitari in H, continua rispetto alla topologia debole degli operatori e tale che $V(s+t) = V(s)V(t)$ per ogni $s, t \in \mathbf{R}$. Una tale famiglia si chiama gruppo a un parametro. Il teorema di Stone afferma che ogni gruppo a un parametro è del tipo $\{e^{itA}\}$ dove A è un operatore autoaggiunto. Questo teorema giustifica il fatto che, nei modelli della meccanica quantistica, l'evoluzione temporale – un gruppo a un parametro di operatori unitari nello spazio di Hilbert degli stati – sia spesso data dall'operatore di Hamilton (A nel precedente capoverso).

2.3. Rappresentazioni lineari.

Definizioni generali. Siano G un gruppo, H uno spazio lineare e $\mathrm{Gl}(H)$ il gruppo dei suoi endomorfismi lineari invertibili. Si dice rappresentazione T del gruppo G nello spazio H un omomorfismo di gruppi $T: G \to \mathrm{Gl}(H)$. Siano T_1 e T_2 due rappresentazioni del gruppo G negli spazi H_1 e H_2 rispettivamente. Si dice operatore di intreccio o morfismo $f: T_1 \to T_2$ un'applicazione lineare $f: H_1 \to H_2$ tale che, per ogni $g \in G$, valga la condizione $f \circ T_1(g) = T_2(g) \circ f$. La rappresentazione T è (algebricamente) irriducibile se nel suo spazio non vi sono sottospazi propri invarianti rispetto a tutti gli operatori $T(g)$, $g \in G$. Le rappresentazioni del gruppo G, come oggetti, e gli operatori di intreccio come morfismi, costituiscono la categoria delle rappresentazioni di G. Gli isomorfismi di questa categoria si dicono usualmente equivalenze fra le rappresentazioni corrispondenti. In essa si hanno anche somme dirette e prodotti tensoriali.

Se il gruppo G ha una struttura ulteriore, ad esempio di spazio topologico o di varietà liscia, è naturale considerare quegli spazi H tali che $\mathrm{Gl}(H)$, o una sua parte, sia dotato di analoga struttura, e limitarsi alle rappresentazioni $G \to \mathrm{Gl}(H)$ che conservano questa struttura e ai corrispondenti operatori d'intreccio.

Una situazione abbastanza consueta si ottiene quando H è uno spazio di Hilbert o di Banach, $\mathrm{Gl}(H)$ è il gruppo degli operatori limitati, G è un gruppo topologico. Le condizioni sulla continuità di T possono essere di tipo diverso, poiché in un gruppo di operatori si possono introdurre piú topologie.

Rappresentazioni di dimensione finita d'un gruppo ciclico. Sia $G = \{g^n \mid n \in \mathbf{Z}\}$, dove g è un generatore del gruppo G. Ogni rappresentazione T del gruppo G nello spazio H è univocamente determinata dall'operatore $A = T(g) \in H$. Se G ha ordine infinito, A può essere qualsiasi; se G ha ordine finito d, allora A deve soddisfare la condizione $A^d = \mathrm{id}$. Due rappresentazioni nello stesso spazio H, corrispondenti agli operatori A_1 e A_2, sono equivalenti se e solo se gli operatori sono coniugati, cioè se $A_1 = BA_2B^{-1}$ per qualche $B \in \mathrm{Gl}(H)$. Pertanto la classificazione delle rappresentazioni di un gruppo ciclico infinito coincide con la classificazione degli operatori, a meno di coniugio. Si è già visto che il principale invariante di un operatore, in questa classificazione, è il suo spettro. S'introducono pertanto alcuni concetti di teoria delle rappresentazioni che a questo sono connessi. Sia H uno spazio a dimensione finita sul campo $\mathbf{C}$.

a) *Operatori diagonalizzabili.* Se A è diagonalizzabile, in H esiste una base di autovettori di A. Ciò significa che la rappresentazione $T: G \to \mathrm{Gl}(H)$ si decompone nella somma diretta di sottospazi unidimensionali. Soltanto la decomposizione in somma $H = \bigoplus_{\lambda \in \sigma(A)} H_\lambda$, dove $H_\lambda = \{h \in H \mid A(h) = \lambda h\}$, $\dim H_\lambda =$ = molteplicità di λ come radice di $\det(A - x \cdot \mathrm{id}) = 0$, ha significato invariante. Per ogni sottorappresentazione T_1, cioè per un sottospazio $H_1 \subset H$ con la proprietà $A(H_1) \subset H_1$, esiste una sottorappresentazione T_2 (cioè un sottospazio A-invariante $H_2 \subset H$) tale che $T = T_1 \oplus T_2$ (cioè $H = H_1 \oplus H_2$ e l'azione $T(g)$ è somma diretta delle azioni). Questa proprietà molto importante si dice riducibilità completa della rappresentazione T.

b) *Operatori non diagonalizzabili.* Un esempio di tale operatore è dato da

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} = \lambda \cdot \mathrm{id} + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

o piú in generale da un blocco di Jordan

$$B_n(\lambda) = \lambda \cdot \mathrm{id} + \begin{pmatrix} 0 & 1 & 0 & \dots & \\ 0 & 0 & 1 & \dots & \\ \dots & & & & 1 \\ 0 & 0 & \dots & & 0 \end{pmatrix}$$

di dimensione $n \times n$, $n \geq 2$. Sebbene $B_n(\lambda)$ abbia un solo autovalore λ, tutto lo spazio H su cui agisce $B_n(\lambda)$ non è un autospazio per B_n, per $n \geq 2$. In effetti

$\dim \operatorname{Ker}(B_n(\lambda)-\lambda\cdot \mathrm{id})=1$ e il corrispondente autospazio non ha un completamento invariante. In tal modo una rappresentazione di un gruppo ciclico con $T(g)=B_n(\lambda)$ non è completamente riducibile.

Si può dimostrare che ogni rappresentazione di dimensione finita di un gruppo ciclico su un campo algebricamente chiuso è equivalente a una somma diretta di rappresentazioni corrispondenti ad alcuni blocchi $B_{n_i}(\lambda_i)$, dove $\{\lambda_i\}$ è lo spettro dell'operatore del generatore $T(g)$.

Su un campo non chiuso la classificazione si complica per il fatto che i punti dello spettro possono non appartenere al campo. La classificazione delle rappresentazioni di gruppo su un campo non chiuso richiede in maniera essenziale che si tenga conto dell'azione del gruppo di Galois della chiusura del campo delle costanti su cui è definito lo spazio della rappresentazione. Questa osservazione è generale e banale. A un livello superiore non è banale il fatto che la classificazione di determinate rappresentazioni di gruppi, quali il gruppo delle matrici su campi «aritmetici» (numeri razionali, numeri p-adici, ecc.), è un problema *duale* a quello della classificazione delle rappresentazioni del gruppo di Galois corrispondente.

Nel seguito, salvo avviso contrario, si avrà a che fare soprattutto con rappresentazioni su $\mathbf{C}$.

c) *Meccanismi di semisemplicità.* Si supponga di voler dimostrare la semisemplicità di una data rappresentazione T del gruppo G (non necessariamente ciclico). Ecco una tecnica standard: T è semisemplice se conserva un prodotto scalare non degenere simmetrico (o hermitiano) nello spazio delle rappresentazioni di H, vale a dire se $(T(g)h_1, T(g)h_2)=(h_1,h_2)$ per ogni $g\in G$, $h_i\in H$. In effetti in questo caso è invariante il completamento ortogonale rispetto a un qualunque sottospazio invariante.

L'esistenza di un tale prodotto scalare viene spesso assunta. Rispetto ad esso tutti gli operatori $T(g)$ saranno ortogonali oppure unitari; da ciò prendono nome le rappresentazioni.

A volte se ne può dimostrare l'esistenza. Ecco un metodo universale di media: se il gruppo G è finito, allora, qualunque sia il prodotto scalare $f(h_1,h_2)$ in H, se ne può costruire uno invariante ponendo

$$(h_1,h_2)=\frac{1}{|G|}\sum_{g\in G} f(T(g)h_1, T(g)h_2).$$

Naturalmente il caso non degenere va trattato in maniera separata.

In tal modo è possibile dimostrare il classico teorema di Maschke: ogni rappresentazione di un gruppo finito G su $\mathbf{C}$ (o su un qualunque campo di caratteristica che non divide l'ordine di G) è completamente riducibile.

Nella categoria dei gruppi topologici compatti la media può essere sostituita con l'integrale invariante di Haar. Grazie a ciò è possibile stabilire che ogni rappresentazione fortemente continua di un gruppo compatto G nello spazio di Hilbert H ammette un prodotto scalare invariante, la cui norma definisce in H la stessa topologia di partenza. Di conseguenza si dimostra che ogni tale rappresentazione si decompone in somma diretta (hilbertiana) di addendi

irriducibili a dimensione finita. (La dimensione finita delle rappresentazioni irriducibili di un gruppo finito segue direttamente dal fatto che ogni vettore $h\in H$ è contenuto nel sottospazio invariante $\sum_{g\in G} T(g)h$; questa osservazione presuppone che si possa utilizzare l'integrazione-media nella dimostrazione ed anche in generale, ma naturalmente ciò non è affatto banale).

d) *Caratteri.* Come si è già visto, una rappresentazione di un gruppo ciclico è semisemplice se e solo se gli operatori sono diagonalizzabili. Inoltre per ritrovare la rappresentazione, a meno di equivalenze, in questo caso è sufficiente conoscere gli spettri degli operatori $T(g)$ (o anche di uno solo di essi che sia generatore).

Ma lo spettro di un operatore lineare a dimensione finita è univocamente determinato quando sono note le tracce di tutti i suoi ordini: in effetti allora, mediante la formula di Newton, si possono ritrovare le funzioni simmetriche elementari degli elementi dello spettro e in seguito questi stessi elementi come radici del polinomio caratteristico. In tal modo si ottiene il seguente risultato.

Si dice carattere χ d'una rappresentazione di dimensione finita T del gruppo G la funzione $\chi: G\to\mathbf{C}$: $\chi(g)=\operatorname{Tr}(T(g))$. Se G è ciclico e T è completamente riducibile, allora T è univocamente determinata dal suo carattere, a meno di equivalenza.

Entrambi questi risultati sono veri per le rappresentazioni di tutti i gruppi finiti, e anche per le rappresentazioni unitarie dei gruppi compatti. Valgono sia la riducibilità completa di tutte queste rappresentazioni che la possibilità di ritrovare ogni rappresentazione a partire dal suo carattere.

2.4. Alcuni risultati classificatori della teoria delle rappresentazioni.

Generalità. Sia G un gruppo topologico compatto, $\hat{G}$ lo spazio delle sue rappresentazioni unitarie irriducibili (si veda l'articolo «Dualità» in questa stessa *Enciclopedia*). Per ogni classe di rappresentazioni (unitarie, complesse) isomorfe $\pi\in\hat{G}$ sia scelto un suo rappresentante $\pi: G\to \mathrm{U}(H_\pi)$ (gruppo degli operatori unitari nello spazio di Hilbert H; π è continua rispetto alla topologia debole). Lo spazio H_π ha dimensione finita.

La misura invariante di Haar su G permette d'introdurre lo spazio $L^2(G)$ delle funzioni a quadrato integrabile. È lo spazio di due rappresentazioni di G: grazie alle traslazioni sinistre e destre, queste rappresentazioni si dicono regolari. Il teorema fondamentale di Peter-Weyl afferma che si ha una equivalenza di rappresentazioni $L^2(G)\simeq \bigoplus_{\pi\in\hat{G}} H_\pi\otimes H_\pi^*$, inoltre l'azione delle traslazioni sinistre (destre) sulle funzioni corrisponde all'azione di G su H_π (su H_π^*).

In particolare, se G è un gruppo finito, $L^2(G)$ può essere identificato con l'anello di gruppo $\mathbf{C}[G]$. Ogni rappresentazione irriducibile di G è equivalente a una rappresentazione unitaria, e pertanto è contenuta in una regolare (di molteplicità uguale alla sua dimensione, o ordine).

Sorgono i seguenti problemi: come descrivere l'insieme $\hat{G}$? Come descri-

vere le realizzazioni concrete di H_π? Come descrivere la categoria delle rappresentazioni (col suo prodotto tensoriale strutturale e le somme dirette) di cui $\{H_\pi\}$ sono i generatori? Quali sono le proprietà funtoriali di questa categoria al variare di G, ad esempio per restrizione al sottogruppo $H \subset G$ o per induzione dal sottogruppo (questi due funtori sono aggiunti; si veda il già citato articolo «Dualità»).

Le risposte a questi problemi sono connesse fra loro e non sono univoche; in particolare risultano utili alcune realizzazioni concrete e lo studio delle connessioni fra queste. Seguono qui alcune informazioni sulle rappresentazioni dei gruppi finiti G.

Prima di tutto, in $L^2(G) = \mathrm{Map}(G, \mathbf{C})$ si introduce il prodotto scalare $(\varphi, \psi) = \sum_{g \in G} \overline{\varphi(g)} \psi(g) / |G|$. Ogni carattere della rappresentazione di dimensione finita T è una funzione centrale su G, vale a dire $\chi(ghg^{-1}) = \chi(h)$ per ogni $g, h \in G$. Si denoti con H il sottospazio di tutte le funzioni centrali in $L^2(G)$. Allora i caratteri $\{\chi_\pi \mid \pi \in \hat{G}\}$ costituiscono una base ortonormale dello spazio H. Da qui segue, in particolare, che l'ordine di $\hat{G}$ è uguale al numero di classi di elementi coniugati nel gruppo G (h_1 e h_2 sono coniugati se $h_2 = gh_1g^{-1}$ per un g opportuno). Se G è abeliano, allora $\hat{G} \simeq G^* = \mathrm{Hom}(G, \mathbf{C}^*) = \{\chi_\pi \mid \pi \in \hat{G}\}$ (si veda «Dualità»). Inoltre è noto che l'ordine delle rappresentazioni irriducibili del gruppo G divide l'ordine di G (e anche l'ordine dei gruppi quoziente di G rispetto a un divisore abeliano normale). Insieme a un corollario del teorema sulla rappresentazione regolare: $\sum_{\pi \in \hat{G}} \chi_\pi(\mathrm{id})^2 = |G|^2$, questi risultati spesso permettono di stabilire se una famiglia di rappresentazioni irriducibili del gruppo, costruita in qualche modo, è completa.

Sia $F \subset G$ un sottogruppo. La restrizione Res_F^G, o Res, di ogni rappresentazione del gruppo G al sottogruppo F definisce un funtore dalla categoria delle rappresentazioni di G alla categoria delle rappresentazioni di F. Questo funtore è compatibile con le somme dirette e i prodotti tensoriali. Ammette come funtore aggiunto l'induzione Ind_F^G o Ind. La forma concreta di Ind sugli oggetti è la seguente: se F è rappresentabile nello spazio H, allora $\mathrm{Ind}_F^G(H) = \mathbf{C}[G] \otimes_{\mathbf{C}[F]} H$ con azione attraverso il fattore sinistro. L'aggiunzione dei funtori si riflette in coniugio dei caratteri delle rappresentazioni rispetto al prodotto scalare standard in $L^2(G)$ e $L^2(H)$. Precisamente, siano χ e φ i caratteri di due rappresentazioni dei gruppi G e F rispettivamente. Allora $(\chi, \mathrm{Ind}\,\varphi)_G = (\mathrm{Res}\,\chi, \varphi)_H$ (formula di reciprocità di Frobenius). In particolare se χ e φ sono caratteri di rappresentazioni irriducibili, allora le molteplicità d'inclusione di χ in $\mathrm{Ind}\,\varphi$ e di φ in $\mathrm{Res}\,\chi$ coincidono.

L'induzione viene spesso usata nella costruzione dei caratteri e delle rappresentazioni di un gruppo piú ampio, partendo dall'insieme delle rappresentazioni di un gruppo piú piccolo, in particolare abeliano. Ad esempio, in accordo col teorema di Artin, i caratteri indotti dalle rappresentazioni di tutti i possibili sottogruppi ciclici di un gruppo finito generano lo spazio delle funzioni

centrali su $\mathbf{Q}$. La restrizione a un gruppo piú piccolo, spesso ancora abeliano, fornisce un utile metodo di classificazione delle rappresentazioni.

Piú avanti si descriverà brevemente come queste idee agiscono nel caso dei gruppi connessi e compatti, in particolare dei gruppi di Lie. Prima comunque si considererà il caso dei gruppi localmente compatti, ma non compatti, che soddisfino una condizione tecnica importante: sono di classe 1. Questa condizione è soddisfatta, in particolare, da tutti i gruppi classici di matrici, come $\mathrm{Sl}(n, \mathbf{R})$, $\mathrm{Sl}(n, \mathbf{C})$, $\mathrm{SO}(p, q)$, ecc. In questo caso il ruolo del teorema di Peter-Weyl viene assunto dalla formula astratta di Plancherel: su $\hat{G}$ esiste una misura $d\mu(\pi)$ tale che $L^2(G)$ si decomponga come integrale diretto di spazi di Hilbert, in generale di dimensione infinita, $\int_{\pi \in \hat{G}} H_\pi \hat{\otimes} H_\pi^* d\mu(\pi)$. Nel caso piú semplice $G = \mathbf{R}$, si ha $\hat{G} = \mathbf{R}$ e questa formula riassume l'essenza della classica analisi di Fourier sulla retta: ai caratteri χ_a, $a \in \hat{\mathbf{R}}$, $\chi_a(x) = e^{iax}$ corrispondono spazi unidimensionali che appartengono a $L^2(\mathbf{R})$ solo «infinitesimalmente». Fin dal suo inizio la teoria dei caratteri richiede l'introduzione di funzioni generalizzate, poiché tali sono le tracce degli operatori unitari a dimensione infinita.

2.5. Classificazione dei gruppi compatti di Lie.

In accordo con il teorema di Hurwitz, $\mathbf{R}$ (numeri reali), $\mathbf{C}$ (numeri complessi) e $\mathbf{H}$ (quaternioni) esauriscono i campi topologici su $\mathbf{R}$. In ognuno di essi si ha un'operazione di coniugio (l'identità in $\mathbf{R}$) che permette d'introdurre il concetto di forma quadratica (hermitiana, hermitiana per quaternioni) definita positiva sugli spazi a dimensione finita K^n. (Rispetto a una base opportuna, questa forma è del tipo $\sum x_i \bar{x}_i$). Il gruppo delle trasformazioni K-lineari di K^n che hanno determinante unitario e che conservano questa forma si denota rispettivamente $\mathrm{SO}(n)$, $\mathrm{SU}(n)$, $\mathrm{Sp}(n)$. Questi gruppi di tre tipi di rotazioni ortogonali, detti gruppi classici di Lie, costituiscono la fondamentale dotazione dei gruppi compatti di Lie e tutti i rimanenti sono loro «parenti prossimi». Precisamente: *a*) con esclusione di alcuni fra i primi valori di n, questi gruppi sono semplici, a due a due non isomorfi, e hanno gruppi fondamentali finiti; *b*) tutti i gruppi di Lie semplici e compatti sono esauriti dalle loro coperture finite, ed anche da una collezione finita di gruppi particolari; questi ultimi hanno realizzazioni canoniche sotto forma di sottogruppi chiusi dei gruppi classici.

Una dettagliata teoria classificatoria, di cui si parlerà piú avanti, completa questo semplice quadro in due aspetti essenziali: in primo luogo è data una caratterizzazione assiomatica di quella naturale classe di gruppi a cui appartengono le serie classiche, e si mettono in luce alcuni elementi particolari di questa classe, che da un punto di vista intuitivo non sono affatto evidenti; in secondo luogo si elabora una tecnica per lo studio strutturale di questi gruppi e delle loro rappresentazioni, la quale ha un particolare valore in tutte le applicazioni: la tecnica delle radici e delle decomposizioni radicali.

Toro. Il piú semplice gruppo compatto e connesso di Lie è la circonferenza $\{z \in \mathbf{C} \mid |z| = 1\}$. Il prodotto diretto di n circonferenze si dice toro n-dimensionale. Ogni gruppo di Lie abeliano, compatto e connesso, è un toro. La teoria strutturale dei gruppi di Lie semplici e compatti si basa essenzialmente sullo studio dei tori massimali in essi contenuti. Il motivo di questo fatto si può ritrovare osservando che il gruppo $\mathrm{SO}(2)$ delle rotazioni del piano euclideo è esso stesso un toro unidimensionale, e che alle decomposizioni di $\mathbf{R}^{2n}$ in somma diretta di piani ortogonali corrisponde un toro n-dimensionale in $\mathrm{SO}(2n)$, relativo alle rotazioni indipendenti di questi piani. In tal modo, anche non sapendo a priori che un dato gruppo è il gruppo delle rotazioni ortogonali da qualche parte, i tori permettono di catturare le «piú semplici rotazioni».

Algebra di Lie. Ogni gruppo di Lie porta con sé anche lo spazio lineare sul quale si rappresenta: si tratta dello spazio tangente nell'unità di $L(G)$, e la rappresentazione $\mathrm{Ad}: G \to \mathrm{Aut}\, L(G)$, detta aggiunta, è tangente nell'unità all'applicazione di coniugio $x \mapsto gxg^{-1}$, $x, g \in G$. Lo spazio $L(G)$ è anche un'algebra di Lie: se con ad si denota l'applicazione tangente a Ad, l'operazione $[x,y] = \mathrm{ad}\, x(y)$ è una parentesi di Lie.

Rappresentazione del toro su un'algebra di Lie. Sia $T \subset G$ un toro. La rappresentazione Ad del toro T su $L(G)$ si decompone nella somma diretta $L_0 \oplus \sum_{i=1}^{m} L_i$, dove la rappresentazione di T su L_0 è banale e ogni L_i è un piano euclideo su cui l'azione di $t \in T$ è la rotazione di angolo $2\pi\vartheta_i(t)$. La struttura euclidea invariante su ogni L_i è univocamente determinata, a meno di un fattore scalare, di modo che i $\vartheta_i(t)$ sono definiti mod $\mathbf{Z}$, ed anche a meno di cambiamenti di segno, corrispondenti a cambiamenti di orientazione su L_i.

Radici. Ogni toro di G è contenuto in un toro massimale; ogni sottogruppo abeliano connesso, contenente un toro massimale, coincide con esso. Se T è massimale, allora il sottospazio $L_0 \subset L(G)$ definito sopra, su cui $L(T)$ agisce in modo banale, coincide con $L(T)$ (nel caso generale $L_0 \subset L(T)$). Sia inoltre T massimale.

L'applicazione $\exp: L(T) \to T$ è un omomorfismo di gruppi topologici $\mathbf{R}^r \to T$, dove $r = \dim T$; il suo nucleo è un reticolo, cioè un sottogruppo discreto di rango r in $L(T)$. Le funzioni $\vartheta_i: T \to \mathbf{R}/\mathbf{Z}$ si possono estendere univocamente a funzioni, ancora denotate con $\vartheta_i: L(T) \to \mathbf{R}$. Si tratta di forme lineari che assumono valori interi sul reticolo di $L(T)$ che è il nucleo di exp.

La famiglia delle forme $\{\pm\vartheta_i\}$ si dice sistema di radici $\Sigma(G)$ del gruppo G. Esso non dipende in maniera essenziale dalla scelta del toro massimale: tutti questi sono a due a due coniugati, e il coniugio porta un sistema di radici nell'altro. Ad esempio nel gruppo $\mathrm{U}(n)$ il toro massimale T è l'insieme delle matrici diagonali $\mathrm{diag}(\exp(2\pi i x_1), \ldots, \exp(2\pi i x_n))$, e le radici sono le differenze $x_r - x_s$, $r \neq s$.

Sistemi astratti di radici. Il sistema delle radici di un gruppo compatto di Lie G è un insieme finito di vettori nello spazio euclideo $L(T)^*$ (con metrica invariante rispetto alle rappresentazioni aggiunte). Questo insieme finito possiede notevoli proprietà di simmetria, che è comodo assiomatizzare introducendo la seguente definizione di sistema astratto di radici Σ nello spazio euclideo E:

a) $\Sigma \subset E$ è un sottoinsieme finito non contenente lo zero e tale che, se $\alpha \in \Sigma$ e $c\alpha \in \Sigma$, allora $c = \pm 1$.
b) Sia $\alpha \in \Sigma$ e $P_\alpha = \{\gamma \in E \mid (\gamma, \alpha) = 0\}$ sia l'ipersuperficie ortogonale ad α. Allora la riflessione rispetto ad essa porta Σ in sé.
c) Si ponga $\{\alpha, \beta\} = 2(\alpha, \beta)/(\beta, \beta)$. Allora $\{\alpha, \beta\} \in \mathbf{Z}$ per ogni $\alpha, \beta \in \mathbf{Z}$.

Ogni sistema di radici di un gruppo compatto di Lie soddisfa queste condizioni. Inoltre nello spazio $E = L(T)^*$ a cui appartengono le radici si isola un reticolo di pesi $\Lambda(G)$ delle funzioni che assumono valori interi nel nucleo di exp in $L(T)$. Tale reticolo è connesso con il sistema di radici come segue:

$$\Lambda_0 = \sum_{\alpha \in \Sigma} \mathbf{Z}_\alpha \subseteq \Lambda(G)$$
$$\Lambda_1 = \{\lambda \in E \mid \{\lambda, \alpha\} \in \mathbf{Z} \text{ per ogni } \alpha \in \Sigma\} \supseteq \Lambda(G).$$

La teoria della classificazione dei gruppi compatti di Lie si sviluppa ulteriormente in due parti: dapprima si dimostra che un gruppo è definito dal suo sistema di radici e dal reticolo dei pesi; in secondo luogo queste coppie vengono classificate.

Gruppi e sistemi di radici. Sia $\Sigma \subset E$ un arbitrario sistema di radici. A partire da Σ si costruiscano due reticoli $\Lambda_0 \subset \Lambda_1$ come prima e si scelga un qualunque reticolo intermedio di rango massimale $\Lambda_0 \subset \Lambda \subset \Lambda_1$. Allora esistono un gruppo di Lie G, connesso e compatto, un toro massimale in esso $T \subset G$ e un isomorfismo $\varphi: E \to L(T)^*$ che porta Σ in $\Sigma(G)$ e Λ in $\Lambda(G)$.

La costruzione di (G, T) è funtoriale rispetto agli isomorfismi di terne (E, Σ, Λ): gli isomorfismi sono indotti dalle applicazioni lineari di E che portano Σ in Σ' e conservano l'operazione $\{\alpha, \beta\}$ (non il prodotto scalare, che non è univocamente definito in $L(T)$).

Classificazione dei sistemi di radici. Se (E_1, Σ_1) e (E_2, Σ_2) sono due sistemi di radici, allora anche $(E_1 \oplus E_2, \Sigma_1 \cup \Sigma_2)$ è un sistema di radici. I sistemi che non si possono rappresentare in questa forma sono detti irriducibili. Ogni sistema di radici si decompone univocamente come somma diretta di componenti irriducibili.

I sistemi irriducibili si possono classificare come segue. Ogni ipersuperficie di E divide Σ in due parti Σ_+ e Σ_- in modo corrispondente alla decomposizione in componenti del complemento dell'ipersuperficie. Si dicono radici positive gli elementi di Σ_+. Si distingue in Σ_+ il sottoinsieme delle radici semplici Π, cioè tali che $\alpha \neq \beta + \gamma$ per ogni $\beta, \gamma \in \Sigma_+$.

Il sistema di radici Σ si ritrova a partire da un oggetto piú semplice, formato da Π e da tutti i numeri $\{\alpha, \beta\}$, $\alpha, \beta \in \Pi$: ogni biiezione $\Pi \to \Pi'$ che conserva questi numeri si estende a un isomorfismo fra i corrispondenti sistemi di radici.

Infine il numero $\{\alpha, \beta\}$ è determinato dall'angolo fra α e β e dalla conoscenza del vettore di modulo maggiore. Dalle proprietà del sistema di radici si può dedurre che, se $\alpha \neq \pm\beta$, gli angoli possono assumere soltanto i seguenti valori: (0) 90^0; (1) 60^0 oppure 120^0; allora $|\alpha| = |\beta|$; (2) 45^0 oppure 135^0; $|\alpha|/|\beta| = \sqrt{2}$; (3) 30^0 oppure 150^0; allora $|\alpha|/|\beta| = \sqrt{3}$.

L'oggetto combinatorio $(\Pi, \{\alpha, \beta\})$ è utilizzato nelle rappresentazioni con diagrammi di Dynkin, nelle quali si presenta come un grafo con le seguenti proprietà: i vertici rappresentano gli elementi di Π; due vertici sono uniti da i spigoli se i è il numero del punto che nel capoverso precedente descrive l'angolo; se $|\alpha| > |\beta|$, sullo spigolo corrispondente si traccia una freccia diretta verso β.

Dopo questo teorema di classificazione dei sistemi di radici si può affermare che la figura 5 riportata nell'articolo «Invariante», VII, p. 929, di questa stessa *Enciclopedia* mostra tutti i diagrammi di Dynkin dei sistemi irriducibili di radici.

Il legame con i gruppi classici è il seguente: $A_n = \Sigma(\mathrm{SU}(n+1))$, $B_n = \Sigma(\mathrm{SO}(2n+1))$, $C_n = \Sigma(\mathrm{Sp}(n))$, $D_n = \Sigma(\mathrm{SO}(2n))$. [I.M.G. e JU.I.M.].

Gerškovič, F. M.
1979 *Ob odnoj invencii Ioganna Sebast'jana Bacha. (K voprosu o projschoždenii klassičeskoj venskoj sonatnoj formy)*, in «Trudy po znakovym sistemam», XI, pp. 44-70.

Hofstadter, D., e altri
1979 *An Eternal Golden Braid*, Basic Books, New York.

Klein, F.
1872 *Vergleichende Betrachtungen über neuere geometrische Forschungen. Programm zum Eintritt in die philosophische Facultät und den Senat der k. Friedrich-Alexanders-Universität zu Erlangen*, Deichert, Erlangen (trad. it. in «Annali di Matematica», serie II, XVII (1889), pp. 307-49).

Lee, T. D.
1970 *Symmetry Principles in Physics and Finite Quantum Electrodynamics*, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, México.

Weyl, H.
1952 *Symmetry*, Princeton University Press, Princeton N.J. (trad. it. Feltrinelli, Milano 1962).

La nozione, in origine piuttosto vaga, di simmetria come **armonia** di proporzioni, permea le piú varie manifestazioni culturali (cfr. **cultura/culture**), dai canoni della bellezza e della salute del **corpo** alle produzioni delle **arti**, dallo studio del vivente (cfr. **individualità biologica, organismo, vita**) alle immagini (cfr. **immagine**) del **mondo** (cfr. anche **cosmologie, universo**). Nel contesto della **fisica** le diverse concezioni della simmetria sono via via slittate dal piano della **rappresentazione** globale (cfr. **locale/globale**) delle prime forme di pensiero razionale emergenti dal mito (cfr. **mythos/logos**) e delle grandi «architettoniche» di idee della **metafisica**, a modellizzazioni piú

locali (cfr. **modello, teoria/modello**) dei fenomeni della **natura**. La rappresentazione matematica dello **spazio-tempo**, la progressiva messa in luce di «principî di conservazione» (cfr. **conservazione/invarianza**), la sostituzione dell'antica teoria degli **elementi** con quella delle particelle elementari (cfr. **particella**) come premessa per la concezione moderna della **materia** e dell'**energia**, hanno fatto sí che le considerazioni in termini di simmetria svolgessero un ruolo sempre piú rilevante sia nella teoria dei **quanti** sia nella **relatività**: del resto, lo studio delle «rotture della simmetria» è oggi uno strumento essenziale per una teoria unificata (cfr. **uno/molti, unità**) nelle varie interazioni fondamentali (cfr. **forza/campo, interazione**). L'elaborazione entro le **matematiche** si è proficuamente intrecciata a questa evoluzione del **paradigma** fisico, come l'indagine matematica di quel che è **invariante** rispetto a un gruppo di trasformazioni sta a indicare (cfr. anche **applicazioni, dualità, funzioni**). La trattazione delle simmetrie a un notevole livello d'astrazione (cfr. **astratto/concreto**) nelle varie **strutture matematiche** consente infine piú adeguate concettualizzazioni per ogni **sistema** in cui pare rilevante l'idea di generazione ricorsiva (cfr. **ricorsività**) di una struttura complessa (cfr. **semplice/complesso**) per mezzo di un piccolo numero di regole (cfr. **formalizzazione, grammatica, logica**).

·h Press,

·ic

Strutture matematiche

1. *Note introduttive.*

1.1. L'uso comune del termine e lo strutturalismo.

La parola 'struttura' viene ampiamente utilizzata nelle matematiche, per esempio in locuzioni come 'struttura di gruppo' o 'struttura di spazio topologico'. Questo termine appartiene però a un livello talmente elevato che già di per sé riflette nella sostanza un'astrazione metateorica. È bene allora confrontare il suo contenuto con l'idea piú vaga di struttura che si trova invece nelle altre scienze. Lo scopo di queste note introduttive è proprio quello di delineare tale confronto.

Il ruolo dello «strutturalismo», come principio metodologico della scienza del XX secolo, può essere significativamente paragonato a quello dell'«evoluzionismo» nella scienza del XIX secolo. I concetti della coppia struttura/evoluzione formano una delle opposizioni fondamentali entro le concezioni razionalistiche della conoscenza, che è a sua volta strettamente legata all'opposizione spazio-tempo. Semplificando al massimo, si può dire metaforicamente che l'evoluzione del sistema è la sua struttura nel tempo, mentre la struttura è l'evoluzione nello spazio. In realtà la situazione è piú complessa. L'organismo biologico non evolve, si sviluppa soltanto. La specie può evolvere; ma la specie non è il fattore biologico primario, è un'astrazione, oltre tutto un'astrazione che si riferisce alla sfera descrittiva della struttura del mondo vivente del pianeta. Quindi l'evoluzione biologica non è la storia del mondo vivente, ma la storia della sua struttura. A sua volta, la descrizione strutturale, ad esempio della lingua swahili, non è la descrizione di una parte o di tutti i fatti linguistici compiuti in un dato lasso di tempo. Essa include la formulazione di regole che operano in un dato momento, le quali consentono di costruire testi di una lingua accettabili grammaticalmente, di interpretarli, di confrontarli con le situazioni, ecc. La descrizione strutturale fissa piuttosto lo spazio delle possibilità che l'elenco delle loro realizzazioni, e il motivo principale per cui tale azione fissante può in generale essere svolta consiste nel fatto che lo spazio delle possibilità evolve lentamente, è cioè stabile nel tempo. Quando la lingua evolve, questo spazio cambia; il cambiamento degli atti discorsivi si compie molto piú rapidamente e non ammette un confronto diretto con l'evoluzione.

1.2. Struttura ed evoluzione: loro complementarità.

Già questa breve analisi mostra che i concetti di struttura/evoluzione si trovano in tipico rapporto di complementarità alla Bohr (cfr. l'articolo «Dualità» in questa stessa *Enciclopedia*) e, come spesso avviene nelle coppie complementari dei principî metodologici della scienza, intere scuole, e talvolta anche intere

epoche, sono legate, nella loro visione scientifica del mondo, a un'esagerata considerazione di uno dei principî a detrimento dell'altro. Questo eccesso è talvolta esplicitamente teorizzato e i suoi vantaggi si pretendono solennemente confermati da risultati concreti; talvolta esso è inconscio e può essere ricostruito solo dallo storico della scienza. Il XIX secolo ha proclamato in biologia, geologia, linguistica la propria concezione dell'evoluzione in maniera pienamente consapevole e, dopo Darwin, si è occupato molto piú dell'evoluzione delle diverse strutture che delle strutture evolventi. Fra i maggiori pensatori che hanno preparato la svolta verso lo strutturalismo del nostro secolo sono da annoverare il fisico Maxwell e il linguista Saussure.

All'alba dell'atomismo scientifico Maxwell ha posto il problema di come la natura, a livello atomico, non conosca evoluzione, e comunque le concezioni evoluzionistiche non siano in grado di spiegare l'esistenza né la forma delle leggi fisiche oggi note. In particolare Maxwell ha scritto: «Di conseguenza ogni molecola, in tutto l'universo, porta su di sé il marchio di un sistema metrico tanto preciso quanto quello del metro delle Archives Nationales o del doppio cubito regale del tempio di Karnak. Non è possibile costruire alcuna teoria evoluzionistica in grado di rendere conto della somiglianza delle molecole, giacché l'evoluzione sottintende necessariamente un cambiamento continuo, mentre una molecola non è capace né di crescita, né di decadenza, né di generare né di distruggere» [citato in Coulson 1962, p. 746].

D'altra parte la grande scoperta di Maxwell è stata quella di introdurre nella fisica un elemento strutturale di tipo nuovo, il campo. Lo studio dettagliato svolto da Maxwell sulle equazioni del campo elettromagnetico e la loro interpretazione fisica ha aperto la strada alle concezioni di campo della teoria quantistica e della gravitazione. Riunite in una cosmologia teorica, queste concezioni hanno portato a una teoria evoluzionistica del mondo fisico (in particolare degli atomi) di una portata che Maxwell non aveva potuto prevedere. D'altra parte la scoperta dell'invarianza delle equazioni di Maxwell relativamente al gruppo di Poincaré ha condotto successivamente all'unificazione di spazio e tempo in un unico *continuum* quadridimensionale. Nelle costruzioni della fisica teorica ciò ha significato per lo meno la fusione dei concetti di struttura e di evoluzione, cosa mai avvenuta prima.

Lo strutturalismo nelle scienze umane del XX secolo deve a Ferdinand de Saussure l'introduzione di alcuni concetti fondamentali fra cui appunto l'opposizione struttura/evoluzione, che Saussure, nel *Cours de linguistique générale* (1906-11), sulla base di materiale linguistico espone sotto forma della nota coppia oppositiva sincronia/diacronia. Le idee di Saussure sulla sistematicità della lingua sul piano sincronico sono state sottoposte a riproduzione industriale quasi di massa e a standardizzazione divenendo, per questa loro qualità, strumento di lavoro per filologi, linguisti ed etnologi del nostro tempo.

1.3. Un cenno allo strutturalismo etnologico e linguistico.

È opportuno qui notare che, analizzando il concetto di struttura, è possibile individuare fra i suoi innumerevoli impieghi due estremi del significato. Il primo si riferisce al riconoscimento di una qualche unità e reciprocità delle parti che formano il tutto, in applicazione agli oggetti reali della conoscenza o, esprimendosi con maggior cautela, in applicazione agli atti di pensiero che formano l'oggetto della conoscenza. Il secondo appare quando viene applicato ai costrutti linguistici e simbolici che oggettivizzano l'atto del pensare e sono chiamati ad esprimere una concezione teorica degli oggetti della conoscenza.

La non-assolutezza di questa contrapposizione è evidente giacché l'atto del pensare che forma l'oggetto è esso stesso oggettivizzato nella costruzione linguistica, e inoltre la concezione teorica può diventare, e diventa a sua volta, oggetto di conoscenza a un grado successivo. Questa gerarchia di strutture si manifesta ad esempio nell'approccio alle concezioni dello strutturalismo etnologico legato in primo luogo al nome di Lévi-Strauss. La primitiva attività mitica, secondo Lévi-Strauss, rappresenta, in misura considerevole, lo strumento logico per chiarire le fondamentali contrapposizioni binarie, «i dati della realtà», e la loro soluzione dialettica, in particolare grazie a identificazioni diverse, trasposizioni e mediazioni. D'altra parte la stessa concezione di Lévi-Strauss, nel piú ampio contesto dello strutturalismo linguistico e della critica letteraria, può essere sottoposta ad analisi da un teorico della scienza, il quale scoprirà in essa la presenza particolare dei principî costruttivi del testo scientifico che influenzarono nel XX secolo logici, linguisti e matematici.

1.4. Matematica e strutturalismo.

La matematica, infine, ha recato un contributo particolare all'ideologia dello strutturalismo, costituendo il modello (sia per imitazione sia per repulsione) di una disciplina che, mentre è altamente strutturalizzata, crea anche un proprio linguaggio descrittivo, potenzialmente di qualunque strutturalizzazione. Tenuto conto dell'«apertura di significato» delle espressioni matematiche, piú volte sottolineata in altri articoli, cioè della possibilità di un ricco spettro per le interpretazioni concrete, questo linguaggio per descrivere le strutture è diventato un modello quasi universale. L'autocoscienza della matematica si manifesta nella creazione di un modello di se stessa al proprio interno: un'immagine ideale della matematica come gerarchia di linguaggi logici o dei loro referenti all'interno di concezioni di teoria degli insiemi o di altro tipo. La scoperta di tratti comuni in costruzioni matematiche diverse, la loro reciproca correlazione e il loro carattere iterativo hanno prodotto cosí una certa immagine dell'universo matematico. In questa immagine un ruolo di primo piano è svolto dal concetto concretizzato e precisato di «struttura matematica», al quale appunto è dedicato questo articolo.

2. *Le strutture secondo Bourbaki: principî generali.*

2.1. Bourbaki e l'emergenza del concetto di struttura matematica.

L'espressione 'strutture matematiche' è stata introdotta, dotata di contenuto e divulgata dal gruppo di matematici francesi noti sotto lo pseudonimo collettivo di Nicolas Bourbaki. Il concetto di struttura matematica, insieme a un piccolo numero di astrazioni metateoriche di alto livello (come il «metodo assiomatico»), è stato posto da questo gruppo a fondamento dell'attuale soluzione del problema dell'unità delle nozioni matematiche. L'idea di struttura matematica fornisce la chiave per la comprensione e la consapevole rielaborazione architettonica dell'universo matematico; il possesso del metodo assiomatico permette di orientarsi nei suoi spazi infiniti. Per questo, parlando di strutture matematiche, non si può trascurare la descrizione della matematica vista con gli occhi di Bourbaki, benché si possa e si debba tener conto dei cambiamenti di prospettiva.

2.2. Il problema dell'unità delle matematiche.

Il problema dell'unità della matematica non è nuovo; si è posto fin dall'epoca in cui la matematica si è configurata come disciplina autonoma. Percorrendo i tentativi storici di risolvere tale problema, si scopre che le divergenze sono iniziate al momento di definire l'oggetto delle scienze matematiche. Le scienze naturali si occupano del mondo esterno escludendone, per quanto possibile, l'uomo; le scienze umane individuano il posto dell'uomo nel mondo esterno e nella società dei suoi simili; ma quali problemi pone la matematica e qual è il senso delle risposte che essa dà?

Non v'è alcun dubbio che una serie di concetti fondamentali della matematica, quali «numero» e «figura», «derivata» e «integrale», risalgano a rappresentazioni della fisica, ma se i teoremi matematici che vi si riferiscono permettono un'interpretazione fisica diretta, piú spesso numerose interpretazioni fisiche mettono in luce analogie insospettate e affinità tra i vari fenomeni del mondo esterno. Tale constatazione può indurre a considerare la matematica come un linguaggio ausiliario delle scienze naturali (soprattutto della fisica), che vive solo in quanto le scienze naturali gli garantiscono un «senso» e un'«applicazione».

Quasi altrettanto evidente è però anche un'altra particolarità caratteristica della matematica che induce a considerare con maggiore attenzione la sua parentela con le scienze dell'uomo. Neppure il frammento piú piccolo del linguaggio matematico ha una funzione meramente descrittiva, nemmeno nelle applicazioni alle scienze naturali. Le premesse possono avere un'interpretazione sperimentale e una giustificazione; le conclusioni possono essere verificate per via sperimentale o falsificate; ma fra le une e le altre, nel corso della deduzione, che costituisce l'essenza stessa del metodo matematico, avviene qualcosa che non sottostà al confronto diretto con una realtà descrivibile.

Questa deduzione si compie nei casi piú semplici, come si vedrà in seguito nell'esempio dell'aritmetica elementare, all'interno di un sistema simbolico, la cui potenziale applicabilità al mondo esterno è assicurata dall'interpretazione operazionale in termini di enumerazione di oggetti materiali, di aggiunta di una quantità di oggetti materiali a un'altra, ecc. Tuttavia un ben noto paradosso consiste nel fatto che nella maggioranza delle applicazioni reali del simbolismo non ci sono oggetti materiali, né processi fisici che rispondano a queste procedure di enumerazione e di aggiunta, e questi oggetti e processi non possono essere neppure pensati come insospettabili intermediari di un processo fisico concreto.

Per rendere piú chiaro questo pensiero, si consideri un esempio tratto da un testo di scienza naturale. «9. *Produzione di energia di una pila.* Il numero di fissioni prodotte per cm^3 per secondo in una pila è $(f_t nv/\Lambda)(3{,}9/7{,}2)$, dove i fattori sono: la frazione di termici assorbita dall'uranio, la densità dei neutroni, il reciproco della vita di un neutrone termico, e la frazione di neutroni assorbita dall'uranio che ha condotto alle fissioni. Assumendo 200 Mev per fissione, dobbiamo moltiplicare il numero di fissioni/sec cm^3 per 200 Mev/fissione = 0,00032 erg/fissione per ottenere erg/sec cm^3. Nell'espressione che ne deriva appare il prodotto nv, il flusso. Dovremmo però usare il flusso medio $\overline{nv}$, e si può mostrare che per un cubo esso è legato a $n_0 v$, il flusso al centro, dall'equazione $\overline{nv} = n_0 v(8/\pi^3)$. Infine, $kW = n_0 v[(8/\pi^3)V(f_t/\Lambda)(3{,}9 \times 3{,}2/7{,}2)10^{-14}]$. Sostituendo i numeri appropriati, questa espressione vale approssimativamente $(4 \times 10^{-9})\, n_0 v$. Si possono ottenere delle potenze dell'ordine di $2{,}5 \times 10^3$ kW (a Clinton) e questo significa che il flusso al centro è approssimativamente di 10^{+12}. A Hanford i flussi di neutroni al centro delle pile sono approssimativamente di $10^{13} n/\text{sec cm}^2$» [Fermi 1945, pp. 531-32].

Si tratta di una breve descrizione di uno dei calcoli tecnologici legati alla conoscenza dell'energia atomica, simbolo del nostro secolo. La semantica esterna di questo testo è molto complessa: collega le astrazioni profonde della fisica («reciproco della vita di un neutrone termico») con le concrete elaborazioni ingegneristiche. Il contenuto matematico, invece, si riduce quasi all'aritmetica elementare; in particolare, per ottenere il risultato finale occorre semplicemente moltiplicare $8/\pi^3$ per $(3{,}9/7{,}2)3{,}2$, mantenendo due-tre cifre significative del risultato. Eppure proprio in queste operazioni aritmetiche, che costituiscono l'esempio piú semplice di deduzione, si riflette, come in una goccia d'acqua, uno dei misteri della scienza matematica: che rapporto hanno con il reattore atomico le operazioni comuni segnate sul foglio, la moltiplicazione 2×9, il riporto dell'avanzo nella colonna successiva, ecc.? Lo svolgimento di questi conti sul calcolatore rende ancora piú evidente la paradossale assenza di parallelismo fra i processi fisici che avvengono nel corso del calcolo e quei processi fisici il cui risultato o le cui caratteristiche hanno come modello questo calcolo.

Ogni deduzione matematica sottintende la subordinazione di una catena di inferenze alle regole della logica, la spiegazione di costruzioni linguistiche che risultano legittime, in breve l'osservanza di alcune leggi di carattere linguistico. Che rapporto c'è fra queste esigenze di una corretta costruzione del testo e la speranza di una verità delle inferenze rispetto al contenuto?

2.3. Filosofia della matematica e strutture matematiche.

La filosofia della matematica, esterna rispetto alla matematica stessa, presuppone risposte diverse a queste domande e, di conseguenza, soluzioni diverse al problema dell'unità delle conoscenze matematiche. Il platonismo suppone che, come la natura, anche la matematica dell'uomo sia il riflesso prestabilito di un mondo ideale posto al di sopra di entrambi. Il realismo suppone che il cervello umano, in quanto parte della natura, sia portatore dell'impronta complessa e mutevole di un grande mondo, il che spiega anche la concordanza dei nostri processi cognitivi sia all'interno della loro struttura organizzativa sia nel loro rapporto con questo mondo.

Proprio i pensatori che hanno usato piú efficacemente la matematica per conoscere la natura sono inclini a lasciare senza risposta la questione delle cause di questa efficacia. Lo stesso Maxwell scrive: «Chi mi condurrà in quella regione ancor piú segreta e nebbiosa in cui il Pensiero si sposa al Fatto, in cui l'operazione mentale del matematico e l'azione fisica delle molecole appaiono nella loro vera connessione? Non passa forse tale strada attraverso la tana del metafisico, disseminata delle ossa dei ricercatori precedenti, e verso cui tutti gli uomini di scienza provano disgusto?» [citato in Coulson 1962, p. 745].

Il concetto di «struttura matematica» è il risultato del tentativo fatto per raggiungere una meta assai piú modesta di quella che si propone qualunque altra filosofia universale della matematica. Tale meta non consiste nel voler spiegare i successi del pensiero matematicizzato, ma solo nel catalogare e sistematizzare i suoi metodi.

Il gruppo Bourbaki fu costituito da professionisti di prim'ordine; la matematica vi era considerata dall'interno, nella sua viva integrità e nelle sue nuove tendenze di sviluppo. Un ruolo non indifferente nell'impostazione del problema sull'organizzazione delle conoscenze matematiche e nella soluzione proposta fu svolto da considerazioni pedagogiche del tutto ovvie. Quando intraprese la propria attività Bourbaki era giovanissimo. Un giovane matematico si trova di fronte alla necessità psicologica di capire che cosa abbiano in comune, ad esempio, la teoria della somma di variabili casuali indipendenti, che gli viene esposta nel corso di calcolo delle probabilità, e la classificazione delle forme quadratiche del corso di algebra lineare. L'istruzione universitaria con la sua tradizionale suddivisione della matematica in analisi, algebra e geometria ha contribuito, e in gran parte contribuisce tutt'ora, alla formazione di specialisti ristretti, esperti solo nella formalizzazione di una disciplina e capaci di usare un ristretto numero di procedimenti per risolvere problemi simili ad altri già risolti. Lo scienziato che si forma e che desidera ampliare le proprie conoscenze e svolgere i propri impegni professionali con maggior efficacia ha bisogno di una chiara formulazione dei principî generali di costruzione della propria scienza, i quali agiscano da principî sia organizzativi sia creativi.

2.4. Il «paradigma» bourbakista.

Una concezione che soddisfa questa esigenza è stata realizzata negli ultimi quarant'anni proprio dal gruppo Bourbaki. L'inizio della sua attività si colloca negli anni '30; nella cerchia dei fondatori rientravano André Weil, Jean Dieudonné, Claude Chevalley. Questo gruppo ha una composizione variabile, soprattutto per l'accordo di ritirarsi al compimento dei cinquant'anni. Frutto concreto del suo lavoro è il corso in piú volumi sugli *Eléments de mathématique*, uscito a partire dal 1939 e che conta attualmente alcune decine di volumi, incluse le edizioni ampiamente riviste. In minor numero sono stati pubblicati singoli articoli firmati Bourbaki; particolare attenzione merita il famoso lavoro programmatico *L'architecture des mathématiques* [1948], che espone brevemente i principî fondamentali degli autori. Altra straordinaria produzione di questo gruppo sono i lavori del seminario Bourbaki, la cui pubblicazione è iniziata nel 1948. Negli ultimi tempi si tengono tre sessioni di seminari l'anno, in febbraio, in giugno e in novembre; ogni sessione comprende una mezza dozzina circa di conferenze, i cui testi vengono preparati in precedenza e pubblicati successivamente. Attualmente le conferenze pubblicate sono piú di cinquecento. Se gli *Eléments de mathématique* [1939 sgg.] possono essere considerati un corso sistematico universale di scienze matematiche di tipo superiore, che espone i concetti fondamentali e i teoremi in ordine logico e col dovuto grado di generalità, le pubblicazioni del seminario Bourbaki rappresentano la piú ricca cronaca degli ultimi progressi in tutti i campi concreti della matematica. Nello stesso tempo, queste pubblicazioni si differenziano da qualunque rivista matematica ordinaria non specialistica. Poiché i lavori del seminario costituiscono in misura considerevole una selezione ulteriore delle pubblicazioni, la scelta stessa dei temi e la loro formulazione nelle relazioni risultano una chiara espressione del sistema di valori e dell'estetica di Bourbaki. Queste pubblicazioni influenzano l'attività dei matematici forse anche piú degli *Eléments*, a cui, dopo gli anni studenteschi, ci si rivolge come a un manuale informativo. Una rassegna di matematica riflessa nello specchio del seminario Bourbaki è stata pubblicata da Jean Dieudonné [1977].

Si è ritenuto necessario dedicare un certo spazio alla descrizione di particolari storici legati alla comparsa del concetto di «struttura matematica», giacché questo concetto porta in sé il marchio del nostro tempo e non si è ancora tanto cristallizzato da diventare impersonale, né è tanto relegato al passato da presentare un interesse esclusivamente storico. In realtà l'attività del gruppo Bourbaki, vista come processo, ha un interesse straordinario per il sociologo della scienza interessato alla formazione e al primo periodo di vita dei paradigmi scientifici nel senso di Thomas Kuhn (cfr. l'articolo «Paradigma» in questa stessa *Enciclopedia*). In questo esempio sono visibili molti tratti caratteristici di tale processo. Purtroppo l'attività di Bourbaki è trattata dai sociologi della scienza molto meno di quanto meriti; la fonte principale delle informazioni qui riportate sono gli articoli di Halmos [1957] e di Dieudonné [1969]. Halmos di-

pinge a grandi linee il ritratto psicologico collettivo di una forte personalità poliedrica e creativa, l'impazienza e l'acutezza d'ingegno appena attenuata da un vivo senso dell'umorismo. Un ricercatore scientifico deciso a dedicarsi alla descrizione di questo fenomeno eccezionale deve possedere notevoli qualità intellettive e coraggio personale.

2.5. Metodo assiomatico e pratica matematica.

Il concetto di struttura si sviluppa, secondo Bourbaki, sulla base dell'unificazione di due aspetti sostanziali del ragionamento matematico. Il primo è una rigorosa sistematizzazione dell'attività linguistica della matematica che si esprime nella prevalenza del «metodo assiomatico». Come si spiega dettagliatamente nell'articolo «Assioma/postulato» di questa stessa *Enciclopedia*, questo metodo si basa sul fatto che il matematico parte da un sistema fissato di premesse alle quali applica in seguito una successione di passi logici elementari, che costituisce il metodo deduttivo. Anche l'elenco di questi passi viene catalogato. Tuttavia, come rileva Bourbaki, «il metodo delle considerazioni, racchiuso nella costruzione di una catena di sillogismi, è solo un *meccanismo* trasformazionale, applicabile indifferentemente a qualunque tipo di premesse e che, di conseguenza, non può caratterizzare la loro natura. In altri termini, si tratta della *forma* esterna che il matematico dà al proprio pensiero, lo strumento che la rende in grado di collegarsi a altri pensieri e, per cosí dire, il *linguaggio* inerente la matematica, ma nient'altro. Ordinare il vocabolario di questo linguaggio e precisare la sua sintassi significa fare un lavoro molto utile, che costituisce in effetti uno degli aspetti del metodo assiomatico, quello che si può propriamente chiamare *formalismo logico* (oppure, come si dice, "logicista"). Ma, e su questo si insiste, si tratta *solo di un aspetto*, e il meno importante» [1948, ed. 1962 p. 37].

La sistemazione del vocabolario e della sintassi del linguaggio matematico sono state oggetto d'interesse di alcune generazioni di matematici. Le principali tappe di questo lavoro sono legate ai nomi di Boole, Bolzano, Cauchy, Dedekind, Cantor, Peano, Frege, Russell e Whitehead, Hilbert, Bernays, Zermelo, Gödel. L'incompletezza di qualunque formalizzazione rilevata da Gödel, che si esprime nella mancanza d'identità fra deduzione formale e verità, ha portato una schiera di matematici a conclusioni pessimistiche riguardo al senso e al valore dell'astrazione matematica fondamentale: l'idea di infinito attuale su cui poggia il concetto di verità. Di fronte al rifiuto di questa idea una delle possibili concezioni positive consiste proprio nel sostenere le proprietà di «meccanismo trasformazionale» della deduzione, il carattere algoritmico, la computabilità e altre simili «proprietà meccaniche» del ragionamento matematico. Per questo Nicolas Bourbaki, che sostiene il punto di vista opposto, è costretto a insistere – e con buoni argomenti – sul fatto che queste proprietà meccaniche rappresentano solo la parte meno interessante del metodo assiomatico. In realtà l'immagine idealizzata della matematica come processo generativo di tutte le deduzioni formali derivanti da un sistema accettato di assiomi non è semplicemente lontana dalla realtà, ma non coglie un aspetto essenziale della realtà stessa: non

riflette il principio della scelta (o generazione selettiva) delle deduzioni utili e produttive rispetto a un oceano di ragionamenti implausibili, anche se formalmente corretti. Certamente essa non riflette neppure il processo di formazione storica concernente la scelta e il cambiamento degli assiomi accettati o la genesi dei concetti fondamentali nei cui termini questi assiomi sono formulati. Qualunque concezione che tenga conto di queste caratteristiche essenziali presuppone piú o meno chiaramente l'accettazione di un sistema di valori che presiedano alla scelta. Se per le tendenze costruttiviste questi principî si riducono all'ideale affidabilità dei giochi finitistici combinatori, e per quelle intuizionistiche si basano su criteri di convinzione intuitiva non completamente definiti, per Bourbaki è determinante la fiducia del matematico nel valore e nella fondatezza di tutto il patrimonio della matematica classica e del suo moderno sviluppo. Un po' meno preciso è l'atteggiamento di Bourbaki nei confronti del problema del mutuo rapporto fra mondo sperimentale e mondo matematico. Come molti altri pensatori, fra cui i fisici Eugene P. Wigner e Richard Ph. Feynman, Bourbaki constata che per descrivere i fenomeni della natura, in particolare del micromondo, risultano inaspettatamente necessarie le costruzioni matematiche che compaiono come risultato di una logica interna di sviluppo e che non hanno sicuramente una «diretta» origine sperimentale. Bourbaki respinge ogni spiegazione; è fuori dubbio tuttavia che il periodico inserimento di strutture matematiche nel mondo delle scienze naturali, soprattutto della fisica teorica, serve a Bourbaki da conferma dei principî di valore che ha adottato.

L'analisi della matematica classica intrapresa da Bourbaki porta anzitutto a rivedere ciò che conviene considerare come suo sistema di concetti fondamentali (dal punto di vista del loro contenuto si dà per scontato che la struttura logica del linguaggio matematico sia già stata analizzata). Si è già ricordato che nella matematica classica l'idea di «numero reale» appartiene ai concetti fondamentali. La differenza fra l'approccio «logicistico» e l'approccio di Bourbaki si può opportunamente rilevare sull'esempio dell'elaborazione di quest'idea. In maniera breve e concisa, l'analisi logica di un numero reale porta alla sua «aritmetizzazione»: si chiarisce che, in qualità di concetto primitivo, è possibile prendere l'insieme dei numeri naturali col relativo sistema descrittivo di assiomi di Peano; quindi una serie di successive costruzioni insiemistiche conduce dai numeri naturali agli interi, ai razionali, e infine ai reali (per mezzo delle sezioni di Dedekind o del completamento secondo Cantor). Con ciò l'analisi logica termina, se tutti i principî logici e insiemistici esplicitati durante il suo svolgimento sono riconosciuti come veri; la teoria dei numeri reali può essere quindi separata dall'aritmetica e fissata in un corso, come avviene nel classico corso di Landau [1930]. Se però i principî esplicitati non sono riconosciuti validi, possono essere sostituiti da altri; tutto l'edificio delle considerazioni è sottoposto a rielaborazione e conduce a concezioni alternative dei numeri reali, ad esempio secondo Brouwer, Heyting oppure Markov.

L'analisi secondo Bourbaki segue un altro progetto. Si studiano le proprietà dei numeri reali che sono utilizzate ad esempio nelle dimostrazioni dell'analisi classica. Si scopre che queste proprietà si suddividono in modo naturale in tre

gruppi nettamente distinti (si tornerà su questo argomento in seguito, nella descrizione delle «strutture generatrici»):

a) Proprietà algebriche. Ne fanno parte le leggi per la somma e il prodotto: associatività, commutatività, distributività... tutti i numeri rispetto alla somma e tutti i numeri diversi da zero rispetto al prodotto formano un gruppo. In particolare sono conseguenze di queste proprietà tutte le identità «letterali» dell'algebra elementare, tipo il binomio di Newton.

b) Proprietà di ordinamento. I numeri reali sono dotati di una relazione binaria di ordine, relazione che entra in numerose disuguaglianze e maggiorazioni dell'analisi classica.

c) Proprietà topologiche. Su queste si basano la definizione di continuità delle funzioni, tutte le costruzioni del calcolo differenziale legate al processo di passaggio al limite, la proprietà di compattezza di un intervallo chiuso, ecc.

Con ciò è chiaro che *ciascuno* dei tre gruppi di proprietà non è specifico della sola teoria dei numeri reali: ciò che è specifico è soltanto che tutti e tre sono imposti contemporaneamente sul medesimo insieme (cfr. del resto quanto osservato nell'articolo «Calcolo» di questa stessa *Enciclopedia*, II, pp. 393, 396 e 417-18). Si consideri, ad esempio, un insieme astratto con due leggi di composizione, somma e prodotto, che siano commutative, associative, distributive e che costituiscano un gruppo rispetto alla somma in tutto l'insieme e rispetto al prodotto in tutti gli elementi diversi dallo zero del gruppo additivo. Un tale insieme si dice campo o corpo (astratto). Il campo o corpo ha numerose proprietà algebriche comuni ai numeri reali e, in particolare, in esso si verificano le identità dell'algebra elementare. Ma una simile classe di oggetti contiene anche campi del tutto diversi: i numeri razionali, algebrici e complessi, i campi finiti di Galois, i campi di funzioni razionali in uno o piú argomenti, ecc. (si vedano gli articoli «Applicazioni» e «Razionale/algebrico/trascendente» in questa stessa *Enciclopedia*). Se poi il concetto di campo si riduce immediatamente al fatto che su uno stesso insieme vengono contemporaneamente imposte due strutture di gruppo commutativo, è chiaro che piú fondamentale risulta il concetto di gruppo, che in genere deve essere studiato dettagliatamente prima di considerare i campi. Infine, un gruppo è un insieme con una legge di composizione binaria dotato di proprietà particolari (associatività, esistenza dell'unità e dell'elemento inverso), e lo studio dei gruppi dev'essere preceduto dallo studio delle proprietà delle leggi di composizione, che sono dotate di un minor numero di condizioni.

L'utilità di quest'approccio si basa sul fatto che, diminuendo la quantità di proprietà iniziali assunte per gli oggetti, aumenta la quantità di oggetti che rientrano sotto la stessa considerazione e, liberandosi dall'ipnosi delle proprietà particolari di una situazione concreta, si comincia a vedere la profonda generalità e l'analogia fra oggetti matematici apparentemente indipendenti. Si facilita il passaggio di un'idea da un campo all'altro, giacché una considerazione suggerita dalle particolarità di una situazione può essere applicabile a un caso molto

piú generale in cui tuttavia l'intuizione che l'ha generata non trova nutrimento. Ad esempio la conoscenza delle proprietà topologiche dei numeri reali in quanto tali, prescindendo dalla concreta topologia reale, conduce all'idea di chiarire se nell'insieme dei numeri razionali non vi siano altre definizioni di «vicinanza» che possano essere usate per il completamento di questo campo, diverso dal completamento reale. Hensel scoprí queste definizioni, le topologie p-adiche; Ostrowski dimostrò che oltre a quella reale esse esauriscono le topologie naturali; e in seguito a lavori di Chevalley e Artin la considerazione congiunta di tutte le topologie è diventata lo strumento fondamentale dell'aritmetica, considerata per molti secoli il regno del discreto (cfr. del resto quanto osservato in questa *Enciclopedia* negli articoli «Geometria e topologia», VI, pp. 695 sgg.; «Numero», IX, pp. 939 sgg.).

Va sottolineata una particolarità dell'analisi, qui brevemente tratteggiata. Passando dalla considerazione di un *oggetto* concreto – i numeri reali – alla considerazione delle sue «proprietà» isolate, ci si trova di fronte all'alternativa di come intendere queste proprietà: come espressioni puramente linguistiche, o in modo estensionale, come classe di oggetti insiemistici dotati di queste proprietà? Nell'articolo «Insieme» di questa *Enciclopedia* si è cercato di spiegare che in realtà vanno considerati in modo complementare entrambi gli aspetti. Nella matematica contemporanea l'aspetto concettuale si esprime nel linguaggio piú o meno intuitivo della teoria degli insiemi; il formalismo linguistico della teoria degli insiemi precisa le rappresentazioni intuitive e, a loro volta, le costruzioni utili che vengono selezionate all'interno di questo formalismo sono definite da considerazioni intuitive. Per questo le ulteriori considerazioni sulle strutture si condurranno in termini di teoria intuitiva degli insiemi, sottintendendo una loro possibile formalizzazione.

2.6. Strutture fondamentali secondo Bourbaki: un esempio.

La seguente esposizione è da considerarsi preliminare: per ulteriori dettagli si veda il § 3, mentre per i concetti fondamentali qui utilizzati, ma non definiti, si vedano gli articoli «Applicazioni» e «Insieme» in questa stessa *Enciclopedia*. Si segue qui il *Fascicule de résultats* di Bourbaki [1939].

Si ricordi che se è data una famiglia di insiemi $\mathfrak{B}$, è possibile costruire nuovi insiemi a partire da essa, considerando l'insieme delle parti degli elementi della famiglia e il prodotto diretto di elementi della famiglia in un ordine definito. Aggiungendo questi alla famiglia originale si possono ripetere le stesse operazioni e cosí via. Si indichi con $\mathfrak{E}(\mathfrak{B})$ la famiglia massimale ottenuta in tal modo e la si chiami scala d'insiemi di base $\mathfrak{B}$. Ad esempio partendo da una base di tre insiemi $\mathfrak{B} = (M, N, P)$, al primo gradino della scala si ottiene la famiglia di quindici insiemi: M, N, P; $\mathcal{P}(M), \mathcal{P}(N), \mathcal{P}(P)$; $M \times M, N \times N, P \times P$; $M \times N, M \times P, N \times P$; $N \times M, P \times M, P \times N$. Ogni elemento R della scala $\mathfrak{E}(\mathfrak{B})$ è assegnato quando si indica la successione di operazioni fondamentali che occorre applicare agli elementi della base $\mathfrak{B}$ per ottenere R; questa successione si dice schema di costruzione.

L'osservazione fondamentale di Bourbaki consiste nel fatto che gli usuali concetti matematici, come gruppo, insieme ordinato, spazio topologico, ecc. possono essere definiti dai seguenti dati: *a*) una base $\mathfrak{B}$; *b*) un elemento σ di uno degli insiemi N della scala $\mathfrak{E}(\mathfrak{B})$; *c*) alcune condizioni su questo elemento σ: gli «assiomi». Queste condizioni possono essere espresse da una formula nel linguaggio formale della teoria degli insiemi, come anche dall'indicazione di quell'insieme di $\mathfrak{E}(\mathfrak{B})$ di cui σ è un elemento. In questa formula rientrano gli elementi della scala $\mathfrak{E}(\mathfrak{B})$.

In una tale descrizione sono essenziali i seguenti oggetti: il termine che descrive N e la formula che descrive le condizioni su σ; gli elementi della base possono essere considerati indeterminati, o variabili; questo termine e la formula descrivono allora un certo tipo di struttura. L'assegnazione di una base concreta $\mathfrak{B}$ e di un elemento concreto σ corrisponde alla scelta di una struttura concreta di dato tipo.

Si illustra ora quanto detto sull'esempio della struttura «spazio topologico». Le corrispondenti formule e termini saranno descritti nel linguaggio esteso L_1 Set che comprende in particolare le operazioni $\mathcal{P}$ e $\times$, utilizzando le consuete abbreviazioni di scrittura (si veda il già citato articolo «Insieme»). La base è costituita da un insieme M, l'insieme dei punti del futuro spazio topologico. La topologia è data da un insieme di sottoinsiemi aperti di M, e cioè da un sottoinsieme $T \subset \mathcal{P}(M)$ che si può anche considerare come elemento $\sigma \in \mathcal{P}(\mathcal{P}(M))$. L'insieme $N = \mathcal{P}(\mathcal{P}(M))$ appartiene alla scala $\mathfrak{E}(\{M\})$, mentre σ fornisce la struttura richiesta. Per descrivere la formula degli «assiomi di spazio topologico» si scrivano ora nel linguaggio L_1 Set le consuete componenti verbali:

a) M e il sottoinsieme vuoto di M sono aperti: $P_1 : \emptyset \in T \wedge M \in T$;
b) l'intersezione di ciascuna coppia di elementi di T appartiene a T: $P_2 : \forall x \forall y (x \in T \wedge y \in T \Rightarrow x \cap y \in T)$;
c) l'unione di una famiglia qualsiasi di elementi di T appartiene a T: $P_3 : \forall x (x \in \mathcal{P}(T) \Rightarrow \bigcup_{y \in x} y \in T)$.

La formula $\mathcal{T} : P_1 \wedge P_2 \wedge P_3$ è l'assioma di spazio topologico. La coppia (σ, assioma $\mathcal{T}$) definisce il tipo di struttura «spazio topologico».

Tutte le proposizioni derivabili in un linguaggio del tipo L_1 Set dai suoi assiomi e dall'assioma $\mathcal{T}$ appartengono alla teoria degli spazi topologici. Aggiungendo «nuovi assiomi», vale a dire rafforzando $\mathcal{T}$, si ottiene la teoria di una struttura piú ricca. Ad esempio, si può aggiungere ai consueti assiomi di topologia l'assioma di separazione secondo Hausdorff, pervenendo alla teoria degli spazi separati. Oppure è possibile riunire in un insieme due strutture, ad esempio di spazio topologico e di gruppo richiedendo che siano compatibili, nel senso che le operazioni di gruppo siano funzioni continue. Si perviene cosí alla teoria dei gruppi topologici.

Può avvenire che gli assiomi di una data struttura siano contraddittori; in accordo con Gödel ciò equivale al fatto che sia vuota la parte T, definita dagli assiomi di qualche insieme della scala $\mathfrak{E}(\mathfrak{B})$ a cui deve appartenere l'elemento σ che definisce la struttura (per lo meno se la teoria è formalizzata in un linguaggio del prim'ordine). Oppure può avvenire che gli assiomi di una data struttura la determinino univocamente a meno di isomorfismi (si veda piú avanti). Da un punto di vista logico questa possibilità è considerevolmente piú sottile; presuppone una gerarchia di almeno due linguaggi in quanto, secondo Skolem, la teoria del prim'ordine dei numeri naturali, ad esempio, se non è contraddittoria, ammette infiniti modelli non isomorfi, e la scelta di uno di essi può essere effettuata soltanto per mezzo di un linguaggio piú potente (cfr. ancora l'articolo «Insieme»). Senza entrare nel merito di questi problemi va osservato che è possibile attribuire un significato all'affermazione che i numeri interi, i numeri reali, la geometria euclidea classica, ecc. sono univocamente determinati dai loro assiomi. È questo un aspetto caratteristico della matematica classica.

Infine la maggior parte delle strutture fondamentali di Bourbaki ammette piú realizzazioni non isomorfe, perfino nel senso classico della parola: sono tali i gruppi, gli spazi topologici, le varietà differenziabili, ecc. Secondo Bourbaki, lo studio di teorie a piú realizzazioni è l'aspetto piú sorprendente che distingue la matematica contemporanea da quella classica.

2.7. Strutture «generative».

Bourbaki sottolinea che il metodo assiomatico, nel quale l'unità centrale dell'analisi è il concetto di struttura, si presenta come un metodo creativo, in quanto permette di osservare analogie esistenti fra teorie a prima vista relative a oggetti diversi. Negli *Eléments de mathématique* la teoria di ciascuna struttura fondamentale viene sviluppata quanto piú è possibile e in maniera conseguente. Il matematico che riconosce nel proprio oggetto di ricerca delle strutture fondamentali può allora consultare un arsenale già pronto d'informazioni e di metodi di lavoro su queste. Fra tutte le strutture si distinguono tre gruppi maggiori di «strutture generatrici», tali che le altre si possano rappresentare come loro derivate. Ne fanno parte:

a) Le strutture algebriche. Sono strutture connesse con varie leggi di composizione, ossia applicazioni $X \times Y \to Z$, come i gruppi, le algebre di Lie, gli anelli, i campi, ecc. La cristallizzazione delle rappresentazioni relative alle strutture algebriche è legata al graduale riconoscimento degli aspetti comuni ai diversi processi di calcolo. Fra le piú note, le quattro operazioni dell'aritmetica elementare: somma, differenza, prodotto e divisione. Storicamente il successivo ampliamento del concetto di numero – dai naturali, interi, reali ai complessi calcolati in modulo, ecc. – si è accompagnato alla scoperta che la forma dell'apparato di calcolo (operazioni fondamentali e identità) rimane immutata sebbene cambino gli oggetti sui quali si conduce il calcolo. Calcoli con matrici e sostituzioni rivelarono la conservazione di alcune proprietà delle operazioni analoghe al prodotto (associatività, esistenza di unità) e la mancanza di altre (commutatività). L'analisi del grande apparato algebrico di tutta la matematica ha condotto gradualmente alla nascita di astrazioni matematiche di due tipi, che sono state codificate nel XX secolo: il concetto di linguaggio formale e il concetto di leg-

ge di composizione astratta su una famiglia di insiemi. Nella matematica attuale entrambi questi aspetti dell'astrazione sono talvolta inseparabilmente mescolati. Un'accurata analisi dei fondamenti mostra la loro complementarità, sebbene sia possibile ridurre i linguaggi a insiemi con leggi di composizione e gli insiemi con leggi di composizione a modelli di un linguaggio. Bourbaki naturalmente dà la preferenza al punto di vista insiemistico.

La piú semplice legge di composizione è un'applicazione $f: E \times E \to E$ che associa a ogni coppia di elementi (a, b) di un insieme E un terzo elemento $f(a, b) \in E$. Le leggi associative, quelle per cui $(ab)c = a(bc)$, assumono usualmente il nome di «prodotti», dove si è notato $f(a, b) = ab$.

A volte un motivo piú profondo di importanza della proprietà associativa consiste nel fatto che è associativa la composizione insiemistica di applicazioni, cioè il prodotto di morfismi nella categoria degli insiemi. Uno dei piú potenti approcci allo studio delle leggi di composizione non associative sta nella loro riduzione a leggi associative: se ad un elemento $a \in E$ si associa l'applicazione $ad_a: E \to E, ad_a(b) = ab$, allora la composizione di applicazioni ad_a sarà associativa.

Una legge di composizione $E \times E \to E$, relativa a elementi di uno stesso insieme, si dice interna. Sono utili e s'incontrano spesso anche leggi di composizione esterne, ad esempio applicazioni $F \times E \to E$ o anche $F \times E \to G$. Si dice che gli elementi di F agiscono come operatori sull'insieme E. Leggi di operatori entrano nella definizione di strutture fondamentali quali moduli su un anello, spazi vettoriali su un campo, ecc. Il caso in cui l'insieme degli operatori è un gruppo e la composizione in F corrisponde alla composizione delle applicazioni è il modello matematico fondamentale del concetto di simmetria (si veda l'omonimo articolo in questa stessa *Enciclopedia*).

Leggi di composizione lineari, come moduli e spazi lineari, costituiscono i modelli fondamentali delle idee di teoria della misura e del principio di «linearità dei piccoli incrementi» che, insieme alle concezioni topologiche, sono alla base dell'analisi matematica, ossia del calcolo differenziale e integrale. Leggi di composizione che definiscono strutture del tipo dei campi e degli anelli rientrano nella teoria dei corpi di numeri reali e complessi, degli anelli di operatori lineari, nella teoria delle equazioni algebriche (anelli di polinomi e loro estensioni), ecc.

b) Le strutture d'ordine costituiscono la successiva classe di strutture generatrici secondo Bourbaki. In matematica gli esempi piú importanti sono forniti naturalmente dai numeri naturali, dai numeri reali e da qualsiasi insieme di numeri reali. L'archetipo psicologico delle strutture d'ordine è l'idea di tempo, reso discreto oppure continuo. Il motivo fondamentale per comprendere le strutture d'ordine fra quelle generatrici è la scoperta dei numeri ordinali e del metodo di induzione transfinita da parte di Georg Cantor. Bourbaki preferisce utilizzare l'induzione transfinita nella forma del lemma di Zorn, la cui formulazione e dimostrazione fanno a meno dell'introduzione dei numeri ordinali. Si può discutere dei vantaggi e dei difetti di questa opzione. L'elegante costruzione degli insiemi totalmente ordinati, realizzata dalla teoria di Cantor, è di fatto quasi sempre inessenziale per l'induzione transfinita. D'altro lato l'idea di «computazione successiva» degli elementi di un insieme qualsiasi di solito dà maggior rilievo sia all'essenza delle applicazioni sia ai loro centri nevralgici, dal punto di vista dei critici dell'infinito attuale.

Le strutture d'ordine assumono un significato sempre maggiore nei tentativi di applicazione della matematica fuori dalla fisica, cioè in biologia e in sociologia. La concezione fisica di «misura numerica» diventa spesso un letto di Procuste sul quale si dispongono male i massicci dati sociometrici o dello stesso genere. Spesso bastano solo valutazioni ordinali, e non è nemmeno detto che quest'ordine sia sempre totale. L'assegnazione di un numero agli elementi di una successione è allora in larga misura arbitraria e la sua elaborazione mediante formule può portare ad artifici metodologici (cfr. del resto il già citato articolo «Numero», in particolare il § 7).

c) Le strutture topologiche costituiscono infine l'ultima classe di strutture generative. Esse sono legate all'archetipo psicologico della «vicinanza», le cui radici riposano nei meccanismi biologici dell'organizzazione dello spazio percettivo intorno a un centro, coincidente col corpo dell'osservatore. La valutazione della vicinanza del cibo, del ricovero, del nemico è infatti vitale per l'esistenza individuale e la continuazione della specie.

Con lo sviluppo della pratica produttiva le strutture di vicinanza si concretizzano in procedure di calcolo approssimato. Poiché i risultati della misura si esprimono con numeri, la topologia nasce prima di tutto come topologia dei numeri reali. Avvicinandosi a quest'idea, la costruzione contemporanea, cioè il concetto di spazio metrico, dà un insieme dotato di una funzione biargomentale detta «distanza» con proprietà che generalizzano quelle usuali della distanza nello spazio tridimensionale. Questa costruzione è eccezionalmente utile. Se ne osservino le principali diramazioni verso le varie linee di sviluppo della topologia.

Conservando la struttura euclidea infinitesimale e la dimensione finita si perviene alle varietà riemanniane. Separando poi il meccanismo algebrico del prodotto scalare e rinunziando al fatto che esso sia una forma definita positiva, si perviene alle varietà pseudo-riemanniane, in particolare ai modelli contemporanei dello spazio-tempo. Conservando soltanto la struttura lineare e la disuguaglianza triangolare, ma rinunziando alla dimensione finita, si perviene alle idee dell'analisi funzionale. Tuttavia qui non si è ancora rinunziato ai numeri reali.

Piú tardi nacque la concezione di intorno e d'insieme aperto, cioè di struttura topologica nel suo aspetto moderno. Gli assiomi di spazio topologico secondo Hausdorff sono una prosecuzione del «linguaggio ε-δ» dell'analisi classica.

Altri aspetti della topologia sono legati al problema della descrizione discreta delle forme geometriche, per mezzo di invarianti combinatori dei ricoprimenti o dei complessi simpliciali, di classi d'omotopia delle applicazioni, di singolarità, ecc.

2.8. L'«architettura delle matematiche».

Il tipo generale di oggetto matematico, secondo Bourbaki, è un insieme astratto dotato di struttura. L'architettura della matematica è determinata dalla gerarchia, dal confronto, dal concatenamento di strutture, a partire da quelle

generatrici che formano modelli sempre piú complessi; la specializzazione a esempi concreti, eventualmente numerici, è collegata al caso generale, che comprende questi esempi in un contesto ampio talvolta fino all'inverosimile. Ciò è evidente in maniera particolarmente chiara nell'esempio della moderna teoria dei numeri.

È lecito affermare che i principî di Bourbaki risultano essenziali non solo perché gran parte della letteratura matematica degli ultimi decenni porta netto il marchio della loro influenza, ma anche perché nello sviluppo delle scienze naturali si manifestano tendenze analoghe. Il linguaggio della fisica teorica, con la sua costante tendenza all'invarianza (o alla «covarianza») del formalismo, che riflette l'indipendenza della descrizione dall'osservatore e da scelte particolari di unità, di sistemi di riferimento, ecc., soprattutto dopo l'unione dello spazio col tempo in un unico continuo infinito, condusse a considerare le stesse strutture dell'analisi interna dei matematici. Le ricerche di Piaget e della sua scuola sulla psicologia della percezione infantile hanno rivelato la presenza di meccanismi biologici, in un certo senso soggiacenti a «strutture generatrici». Schematicamente, un'analisi sottile, storicamente realizzata molto tardi, al nuovo livello riconduce a concezioni prematematiche del tipo «piú/meno», «vicino/lontano», la cui essenza vitale, non razionalizzata, appartiene al periodo iniziale della vita.

2.9. Due cambiamenti di punto di vista.

Ciò nonostante, qualunque particolarità nella formulazione concreta dei principî di Bourbaki può essere storicamente superata. Vanno sottolineati due cambiamenti di punto di vista, che si sono prodotti dal tempo in cui nacque il progetto degli *Eléments*. Il primo è legato al fatto che Bourbaki sottovalutò le idee di algoritmo, costruibilità, programma. Non c'è dubbio che oggi il concetto di algoritmo, in una delle sue formulazioni moderne, debba essere collegato a numerose strutture matematiche fondamentali e trovare un posto corrispondente negli *Eléments*. Il fatto che ciò non sia avvenuto si spiega con motivazioni storiche e psicologiche. Come si è già ricordato, all'inizio del suo gigantesco lavoro Nicolas Bourbaki era notevolmente giovane. Nello stesso tempo veniva creata la teoria delle funzioni ricorsive; evidentemente non era del tutto chiaro in quale misura questa teoria appartenesse alla matematica e in quale alla metamatematica, con cui i rapporti dei matematici attivi sono sempre stati tutt'altro che univoci. Ritenendo necessario etichettare le convenzioni del formalismo nella grandiosa *Théorie des ensembles* [1954], Bourbaki formulò di fatto il proprio linguaggio di lavoro nel breve *Fascicule de résultats* [1939] e in seguito si disinteressò di queste convenzioni. Per questo una parte essenziale di genuini risultati matematici della metamatematica fu trascurata, inclusa una delle strutture fondamentali.

Il secondo cambiamento di cui sarebbe in linea di principio possibile tener conto in successive edizioni degli *Eléments* è piú recente e legato alla comprensione del ruolo delle categorie (cfr. ancora «Applicazioni»). Va ricordato che il punto di vista categoriale applicato alle strutture di un dato tipo, ad esempio

agli insiemi ordinati, ai gruppi o agli spazi topologici, consiste in questo: prima di tutto si considera la classe di tutte le strutture del dato tipo, i cui elementi sono detti oggetti della categoria. Per rendere quanto è dovuto alle difficoltà insiemistiche, occorre dare senso alla parola 'tutte' nella frase precedente, convenendo ad esempio di lavorare in un universo insiemistico fissato, altrimenti ci si trova pericolosamente vicini al paradosso di Russell (cfr. ancora l'articolo «Insieme»). Occorre quindi definire chiaramente i morfismi della categoria. Nei casi piú semplici la scelta della classe dei morfismi è evidente e univoca: nella categoria dei gruppi sono le applicazioni insiemistiche compatibili con le leggi di gruppo, vale a dire gli omomorfismi di gruppo. Tuttavia questa scelta spesso non si presenta né evidente, né univoca. Fra gli oggetti della categoria degli insiemi ordinati può risultare ragionevole considerare applicazioni strettamente crescenti, oppure crescenti ma non strettamente. Fra spazi topologici si possono considerare come morfismi le arbitrarie applicazioni continue, ma spesso è produttivo considerare come tali le classi di omotopia di applicazioni continue. Il concetto di morfismo fra oggetti con una data struttura non deriva cosí univocamente dalla definizione della struttura, ma dev'essere assegnato dall'esterno.

La scelta della definizione dei morfismi è spesso dettata dalla necessità di adattarla a qualche tipo di costruzione naturale, interessante fra gli oggetti con la data struttura. Ad esempio nella teoria delle varietà algebriche il punto di vista classico era essenzialmente «birazionale», in particolare si consideravano morfismi delle applicazioni che potevano avere singolarità su sottovarietà di dimensione minore. Il cambiamento di accento che si verificò con l'introduzione formale della categoria delle varietà algebriche solo con applicazioni regolari in qualità di morfismi si consolidò definitivamente quando fu chiaro che la costruzione delle coomologie dei fasci coerenti ammette buone proprietà funtoriali proprio per questa ristretta classe di morfismi.

Costruzioni naturali fra oggetti di categorie che si comportavano bene rispetto ai morfismi furono in seguito dette funtori e il punto di vista per il quale le categorie e i funtori appartengono ai concetti matematici fondamentali, e le costruzioni fondamentali devono essere funtoriali, si diffonde sempre piú. Il riconoscimento di ciò è legato ai successi della topologia e in seguito della geometria algebrica: il fiorire rigoglioso di queste scienze, cominciato negli anni '50, ha prodotto una grande influenza sul contenuto e sulla forma della matematica contemporanea. Lo sviluppo di questo periodo è ben documentato nei lavori del seminario Bourbaki, mentre nuove redazioni degli *Eléments*, con i naturali aggiornamenti temporali, riflettono tale cambiamento di punto di vista. Per sua propria essenza esso si accorda perfettamente alle concezioni fondamentali di Bourbaki: la matematica delle categorie rimane insiemistica e la generalità delle considerazioni, che suggerisce analogie fra teorie, è raggiunta in misura ancora maggiore di quando ci si limita alle strutture. Inoltre, come si è tentato di dimostrare negli articoli «Dualità» e «Applicazioni» di questa *Enciclopedia*, i punti di vista categoriale e strutturale si trovano in un rapporto di complementarità reciproca abbastanza complesso: essi non si escludono, ma si arricchiscono l'un l'altro.

2.10. Ancora sul punto di vista categoriale.

Negli ultimi anni il punto di vista categoriale ha cominciato a esercitare una profonda influenza non solo sulla costruzione della matematica complessivamente, ma anche sulla concezione dei suoi fondamenti. In una serie di lavori Lawvere e la sua scuola hanno sintetizzato le idee di alcune tendenze sorte in maniera completamente isolata: dalla logica, in particolare intuizionistica, dalla teoria delle topologie astratte di Grothendieck, dalla teoria delle probabilità e, in una certa misura, dalla fisica teorica. Vale la pena di descrivere brevemente il senso generale di questi lavori.

Nella teoria delle categorie esiste la costruzione generale di «relativizzazione»: se sono dati una categoria C ed un suo oggetto B, si può considerare una nuova categoria C_B «di oggetti relativi», o «oggetti su B». I suoi oggetti sono i morfismi $X \to B$, dove X è un oggetto variabile di C, e i morfismi di $X \to B$ in $Y \to B$ sono costituiti dai morfismi $X \to Y$ che commutano con le applicazioni «strutturali» di X e di Y su B. Se Ens è la categoria degli insiemi, è comodo rappresentare gli oggetti della categoria Ens_B come famiglie di insiemi $X_b = f^{-1}(b)$ parametrizzate dai punti $b \in B$. In questo senso Ens_B arricchisce l'usuale categoria degli insiemi. In tal modo si possono relativizzare tutte le relazioni e le costruzioni di teoria degli insiemi: ad esempio la relazione di inclusione $X \subset Y$ in Ens, che da un punto di vista logico può essere solo verificata oppure non verificata, corrisponde ora a un piú sottile valore di verità: il sottoinsieme $\{b \in B \mid X_b \subset Y_b\} \subset B$ come elemento dell'algebra di Boole dei sottoinsiemi di B. Sono possibili numerose varianti di questa interpretazione. Ad esempio nella categoria degli «spazi topologici su B» è naturale talvolta considerare i morfismi $X \to B$ non come famiglia di fibre, ma come fascio delle sezioni locali di questo morfismo su tutti i possibili insiemi aperti di B. A sua volta questo fascio può essere dotato della struttura di fascio di gruppi, o di spazi topologici, ecc. Dal punto di vista logico queste costruzioni costituiscono un potente metodo per la formazione di modelli non standard di teorie classiche, nelle quali la motivazione «non standard» è intuitivamente ed eccezionalmente chiara: l'inclusione di un insieme di parametri variabili della teoria B. L'interpretazione di B come parametro temporale è possibile nello spirito dell'intuizionismo, con cui si realizza l'inclusione inaspettata del formalismo intuizionista nella matematica classica.

Si riassumerà ora la parte non formale delle considerazioni qui esposte. La dialettica dell'opposizione struttura/evoluzione riflette la generale opposizione scientifica essere/divenire. Il successo della teoria degli insiemi come linguaggio universale della matematica e della concezione delle strutture matematiche come realizzazione concreta di questo linguaggio ha segnato uno spostamento d'accento sui primi elementi di queste opposizioni. Le concezioni relative sempre piú diffuse nei fondamenti, gli assiomi categoriali, ecc. segnano il movimento inverso, quando «divenire» e «incompiuto» sono considerati primitivi. Le parole di Lawvere esprimono chiaramente questo cambiamento di valori: «Ogni nozione di costanza è relativa, essendo derivata percettivamente o concettualmente come caso limite della variazione, e l'indiscusso valore di queste nozioni nello spiegare la variazione è sempre limitato da questa origine. Ciò è vero in particolare per la nozione di insieme costante e spiega cosí perché tanta parte di teoria intuitiva degli insiemi si trasporti in qualche forma nella teoria degli insiemi variabili» [1973, p. 136].

3. *Le strutture secondo Bourbaki: descrizione formale.*

In questo paragrafo si daranno le definizioni fondamentali e le costruzioni legate al concetto di struttura secondo Bourbaki. Negli *Eléments de mathématique* queste definizioni sono esposte soltanto al livello del linguaggio formale della teoria degli insiemi. Per semplificare il passaggio al punto di vista delle categorie e dei topoi, descritto piú avanti, si darà un'interpretazione parallela utilizzando il concetto di universo.

3.1. Il linguaggio formale.

Si consideri fissato un linguaggio formale della teoria degli insiemi, del tipo descritto nel § 2 dell'articolo «Insieme» di questa stessa *Enciclopedia*: L_1 Set. Per economia di scrittura, e per avvicinarla all'usuale standard matematico, si lavorerà su una variante allargata del linguaggio, nel quale si abbiano termini del tipo $\mathcal{P}(x)$ (l'insieme delle parti dell'insieme x); $x \times y$ (prodotto diretto degli insiemi x e y), $\{y \mid P(y, y_1, \ldots, y_r)\}$ (insieme di tutti gl'insiemi y che verificano la condizione P, formulata per mezzo di insiemi «parametrici» ausiliari $y_1, \ldots, y_r$), ecc.

3.2. L'universo.

Si fissi un insieme $\mathfrak{U}$ (i cui elementi sono pure insiemi) soddisfacente le seguenti condizioni:

a) se $x \in \mathfrak{U}$ e $y \in x$, allora $y \in \mathfrak{U}$;
b) se $x, y \in \mathfrak{U}$, allora $\{x, y\} \in \mathfrak{U}$;
c) se $x \in \mathfrak{U}$, allora $\mathcal{P}(x) \in \mathfrak{U}$;
d) se (x_α) è una famiglia di elementi di $\mathfrak{U}$, indiciata dall'insieme $I \in \mathfrak{U}$, allora $\bigcup_{\alpha \in I} x_\alpha \in \mathfrak{U}$.

Un insieme $\mathfrak{U}$ di questo tipo sarà detto universo e l'interpretazione standard del linguaggio L_1 Set sarà considerata in $\mathfrak{U}$. Si consideri l'universo $\mathfrak{U}$ come modello intuitivo della teoria degli insiemi. (Per le costruzioni e una descrizione piú dettagliata degli universi cfr. ancora «Insieme», § 4). Utilizzando una variante intuitiva degli assiomi di Zermelo-Fraenkel (cfr. l'articolo «Assioma/postulato») si può mostrare che qualunque universo $\mathfrak{U}$ è chiuso rispetto alle seguenti costruzioni: passaggio alle parti di un insieme di $\mathfrak{U}$, quoziente rispet-

to a una relazione di equivalenza, formazione delle n-ple ordinate, costruzione di somma e prodotto di famiglie di insiemi di $\mathfrak{U}$ indiciati da elementi di $\mathfrak{U}$, ecc. Inoltre ogni applicazione $x \to y$ dove $x, y \in \mathfrak{U}$ (o piú precisamente il suo grafico, si veda l'articolo «Applicazioni») appartiene a $\mathfrak{U}$. Considerando queste applicazioni come morfismi, con l'usuale composizione insiemistica, è possibile trasformare $\mathfrak{U}$ in una categoria (cfr. ancora «Applicazioni»).

3.3. L'universo come categoria.

La categoria $\mathfrak{U}$ è dotata, oltre che degli usuali funtori $\mathrm{id}: \mathfrak{U} \to \mathfrak{U}$ e Hom, del funtore prodotto diretto e di una particolare applicazione sugli oggetti $\mathcal{P}: \mathfrak{U} \to \mathfrak{U}$ (insieme delle parti). Si può trasformare in un funtore in due modi. All'applicazione $f: X \to Y$ si può associare l'applicazione $f^*: \mathcal{P}(Y) \to \mathcal{P}(X)$ che a ogni parte di Y associa la sua controimmagine in X; in tal modo $\mathcal{P}$ diventa un funtore controvariante. Si può anche associare a ogni parte di X la sua immagine in Y: $f_*: \mathcal{P}(X) \to \mathcal{P}(Y)$. In tal modo $\mathcal{P}$ diventa un funtore covariante. Se f è un isomorfismo (vale a dire un'applicazione biunivoca di insiemi), allora f^* e f_* sono isomorfismi inversi $\mathcal{P}(X) \leftrightarrows \mathcal{P}(Y)$.

Un ruolo importante è svolto anche dall'applicazione $f_!: \mathcal{P}(X) \to \mathcal{P}(Y)$ che associa a ogni parte $Z \subset X$ il sottoinsieme $f_!(Z) \subset Y$ costituito da quegli elementi $y \in Y$ tali che $f^{-1}(y)$ sia interamente contenuto in Z.

3.4. Schemi di costruzione nell'universo.

Dai funtori id, $\times$ e $\mathcal{P}$ (nell'estensione controvariante o covariante) è possibile, per mezzo della composizione, formare nuovi funtori, ad esempio: $\mathcal{P}_0(\mathrm{id} \times \mathrm{id} \times \mathcal{P}): \mathfrak{U} \times \mathfrak{U} \times \mathfrak{U} \to \mathfrak{U}: (x, y, z) \to \mathcal{P}(x \times y \times \mathcal{P}(z))$. Ogni funtore S che si può ottenere in questo modo si dice schema di costruzione di scala. Sostituendo al posto degli argomenti di uno schema di costruzione qualche insieme di $\mathfrak{U}$, si ottiene il risultato dell'applicazione di questo schema di costruzione ai dati insiemi. Sia B un insieme di elementi di $\mathfrak{U}$. Il piú piccolo sottoinsieme contenente B e chiuso rispetto alle applicazioni ai suoi elementi di tutti gli schemi di costruzione si dice scala di insiemi di base B. Piú precisamente questa definizione si può formulare in maniera comoda nei termini di un linguaggio formale.

3.5. Schemi di costruzione e linguaggio formale.

Uno schema di costruzione a m-elementi è assegnato da una successione di coppie di numeri naturali $(a_1, b_1), \ldots, (a_m, b_m)$ soddisfacenti le seguenti condizioni: a) se $b_i = 0$ allora $1 \leq a_i \leq i-1$; b) se $a_i \neq 0$ e $b_i \neq 0$ allora $1 \leq a_i \leq i-1$ e $1 \leq b_i \leq i-1$. Un simile schema di costruzione si applica induttivamente a una successione di n termini $E_1, \ldots, E_n$ del linguaggio, dove $n = \max b_i$ (rispetto alle coppie $(0, b_i)$) e fornisce una successione $A_1, \ldots, A_m$ di termini di questa teoria. Precisamente, se $a_i = 0$ allora $A_i = E_{b_i}$; se $b_i = 0$ allora $A_i = \varphi(E_{a_i})$; se $a_i \neq 0$ e $b_i \neq 0$ allora $A_i = A_{a_i} \times A_{b_i}$. L'ultimo termine A_m è il risultato dell'applicazione dello schema S ai termini di base $E_1, \ldots, E_m$. L'interpretazione del linguaggio formale nell'universo $\mathfrak{U}$ permette di passare da una definizione all'altra.

In altri termini, assegnare uno schema di costruzione non consiste soltanto nell'assegnare un funtore di tipo $\mathcal{P}(\mathcal{P}(X) \times \mathcal{P}(Y))$, ma anche nell'assegnare una successione per la sua costruzione per mezzo di passi elementari. L'inclusione di questa successione nella definizione è comoda per dare costruzioni induttive e dimostrazioni; pertanto uno stesso funtore può essere descritto da schemi diversi. Ad esempio conducono a $\mathcal{P}(\mathcal{P}(X) \times \mathcal{P}(Y))$ gli schemi $(0,1)$, $(0,2)$, $(1,0)$, $(3,0)$, $(2,0)$, $(4,5)$ oppure $(0,2)$, $(0,1)$, $(1,0)$, $(2,0)$, $(4,0)$, $(5,3)$.

3.6. Relazioni trasportabili.

Siano $x_1, \ldots, x_n, s_1, \ldots, s_p$ lettere dell'alfabeto di un linguaggio, $A_1, \ldots, A_m$ termini non contenenti le lettere x_i e s_j $(1 \leq i \leq n,\ 1 \leq j \leq p)$. Si scelgano degli schemi di costruzione a $(n+m)$ termini $S_1, \ldots, S_p$ (nella variante linguistica) e si consideri la relazione (formula) $T(x_1, \ldots, x_n, s_1, \ldots, s_p): s_1 \in S_1(x_1, \ldots, x_n, A_1, \ldots, A_m) \wedge \ldots \wedge s_p \in S_p(x_1, \ldots, x_n, A_1, \ldots, A_m)$. Una simile relazione si dice tipificazione delle lettere $s_1, \ldots, s_p$. Se è assegnata un'interpretazione del linguaggio nell'universo $\mathfrak{U}$, i valori di x_i sono detti insiemi fondamentali elementari e quelli di A_j insiemi ausiliari elementari.

Sia $R(x_1, \ldots, x_n, s_1, \ldots, s_p)$ una relazione (formula) comprendente alcune fra le lettere x_i, s_j e, eventualmente, altre lettere. Si scelgano due interpretazioni del linguaggio nelle quali $A_1, \ldots, A_m$ assumano gli stessi valori (che si indicheranno ancora con A_i), e x_j rispettivamente i valori x_j' e x_j''. Si supponga inoltre che siano date delle biiezioni $f_j: x_j' \to x_j''$. Per la prima interpretazione sia verificata la relazione di tipificazione $T(x_1', \ldots, x_n', s_1', \ldots, s_p')$. Si ponga s_j'' = immagine di s_j' nelle applicazioni $f_i: x_i' \to x_i''$, $\mathrm{id}: A_k \to A_k$ estese a $S_1, \ldots, S_p$ con la definizione covariante di $\mathcal{P}$. Allora sarà verificata anche la relazione di tipificazione $T(x_1'', \ldots, x_n'', s_1'', \ldots, s_p'')$. Se, sotto queste condizioni, le relazioni $R(x_1', \ldots, x_n', s_1', \ldots, s_p')$ e $R(x_1'', \ldots, x_n'', s_1'', \ldots, s_p'')$ sono equivalenti, allora si dice che R è una relazione trasportabile rispetto alla tipificazione T.

L'equivalenza di relazioni si può qui intendere in un senso della teoria formale degli insiemi anche piú forte di quello iniziale. Ad esempio, per la tipificazione di $s_1 \in x_1 \wedge s_2 \in x_1$ la relazione $s_1 = s_2$ è trasportabile, mentre la relazione $x_1 = x_2$ non lo è.

3.7. Tipi di struttura.

Un tipo di struttura è descritto da un testo Σ nel linguaggio della teoria degli insiemi, il quale è costituito dai seguenti sottotesti:

a) una collezione non vuota di lettere $x_1, \ldots, x_n, s$ appartenenti all'alfabeto del linguaggio, fra le quali non vi siano simboli di costante; queste lettere devono essere a due a due diverse; $x_1, \ldots, x_n$ denotano insiemi elementari della struttura;

b) una collezione di termini del linguaggio $A_1, \ldots, A_m$, non contenenti le lettere x_i, s. Questa collezione può essere vuota; i termini denotano insiemi ausiliari della struttura;

c) una tipificazione $T(x_1, \ldots, x_n, s)$ della forma $s \in S(x_1, \ldots, x_n, A_1, \ldots, A_m)$ dove S è uno schema di costruzione a $(n+m)$ termini;

d) una relazione $R(x_1, \ldots, x_n, s)$ trasportabile rispetto alla tipificazione T.

La teoria formale del tipo di struttura Σ è la collezione di formule logicamente derivate dagli assiomi della teoria degli insiemi e dall'assioma ulteriore $T \wedge R$.

Sia data un'interpretazione della teoria formale del tipo di struttura Σ nell'universo intuitivo degli insiemi (o in una teoria formale piú forte). $X_1, \ldots, X_n$, S siano i valori delle lettere $x_1, \ldots, x_n$, s in questa interpretazione (cioè insiemi o costanti di una teoria piú forte) e siano $A'_1, \ldots, A'_m$ i valori dei termini $A_1, \ldots, A_m$. Si dirà allora che S fornisce una struttura di tipo Σ alla famiglia di insiemi elementari $X_1, \ldots, X_n$ e di insiemi ausiliari $A'_1, \ldots, A'_m$. La condizione T è una caratterizzazione tipica del tipo di struttura, mentre la condizione R un assioma del tipo di struttura.

Nel punto precedente è stata definita la tipificazione di una famiglia finita di lettere $s_1, \ldots, s_p$; nella descrizione del tipo di struttura è sempre possibile limitarsi alla tipificazione di una sola lettera, pur di sostituire la famiglia di schemi di costruzione $S_1, \ldots, S_p$ con lo schema di costruzione $S_1 \times \ldots \times S_p$ e $s_1, \ldots, s_p$ con $s = (s_1, \ldots, s_p)$. Analogamente una famiglia finita di assiomi $R_1, \ldots, R_q$ può essere sostituita dall'unico assioma $R_1 \wedge \ldots \wedge R_q$. In una descrizione non formale del tipo di struttura, queste condizioni, come molte altre, non sono necessariamente rispettate.

Nelle note introduttive si è indicata la descrizione del tipo di struttura «spazio topologico». Per chiarire la posizione generale si aggiungono ora due altre descrizioni di tipi di struttura, che fanno parte di quelle generatrici.

3.8. Insieme ordinato.

Si ha qui un insieme elementare x e nessun insieme ausiliario. Il grafico della relazione di ordine è definito dalla tipificazione $s \in \mathcal{P}(x \times x)$. Per formulare gli assiomi occorre utilizzare due operazioni fra corrispondenze: la composizione e l'inversione. Una corrispondenza fra gli insiemi x, y è un sottoinsieme $u \subset x \times y$, cioè un elemento $u \in \mathcal{P}(x \times y)$. Sia $u \subset x \times y$ e $v \subset y \times z$; si pongano $u \circ v$ = proiezione su $x \times z$ del sottoinsieme $(u \times z) \cap (x \times v) \subset x \times y \times z$, e $u^{-1} = \{\langle b, a\rangle \mid \langle a, b\rangle \in u\} \subset y \times x$. (Il lettore può scrivere queste formule nella piú semplice variante del linguaggio L_1 Set). Dopo questi chiarimenti, l'assioma di insieme ordinato può essere scritto nella forma $s \circ s = s \wedge s \cap s^{-1} = \Delta_x$, dove $\Delta_x \subset x \times x$ è la diagonale: $\Delta_x = \{\langle a, a\rangle \mid a \in x\}$. Non è difficile verificare la trasportabilità di questa relazione.

3.9. Gruppo.

Anche qui si ha un insieme elementare x e nessun insieme ausiliario. È comodo dare la tipificazione per una terna di elementi $s = \langle e, m, i\rangle$, dove $e \in x$, $m \in \mathcal{P}(x \times x \times x)$, $i \in \mathcal{P}(x \times x)$ sono rispettivamente l'unità, il grafico della legge di composizione $(a, b) \to ab$ e il grafico dell'applicazione di inversione $a \mapsto a^{-1}$. Gli usuali assiomi di gruppo, associatività, assiomi di unità e dell'elemento inverso, si scrivono senza difficoltà nel linguaggio formale. La loro trasportabilità è oltremodo evidente.

3.10. Strutture derivate.

Partendo da alcuni tipi di struttura assegnati è possibile ottenerne di nuovi per mezzo di una serie di costruzioni spesso utilizzate. Ne fanno parte:

a) unione di piú strutture nello stesso sistema di insiemi fondamentali (gruppo topologico, spazio lineare metrico, ecc.);

b) arricchimento della struttura con l'aggiunta di nuovi assiomi (insiemi totalmente ordinati rispetto a quelli ordinati, gruppi di Lie rispetto ai gruppi topologici, ecc.);

c) processo inverso: impoverimento della struttura (fino a dimenticare tutto e lasciare soltanto la struttura insiemistica fondamentale).

Una parte di questa procedura può essere descritta in maniera formale nella metateoria del linguaggio della teoria degli insiemi, come fa Bourbaki. Non ci si soffermerà su questo argomento.

3.11. Il problema dei morfismi.

Si consideri un tipo di struttura Σ. Per semplicità si supporrà di avere un unico insieme fondamentale e che la tipificazione sia data per un unico elemento della scala. Il concetto di morfismo di strutture di tipo Σ può essere assegnato per mezzo di un termine $\sigma(x, y, s, t)$ in cui le lettere sono a due a due diverse e non costanti. Questo termine deve soddisfare le seguenti condizioni: *a*) dagli assiomi del tipo di struttura Σ per (x, s) e (y, t) segue che il termine $\sigma(x, y, s, t)$ è contenuto nell'insieme delle applicazioni di x in y; *b*) nelle medesime ipotesi, se $f \in \sigma(x, y, s, t)$ e $g \in \sigma(y, z, t, u)$ allora $g \circ f \in \sigma(x, z, s, u)$; *c*) una biiezione $f: x \to y$ è un isomorfismo di strutture se e solo se $f \in \sigma(x, y, s, t)$ e $f^{-1} \in \sigma(y, x, t, s)$. In questa scrittura vi sono alcune imprecisioni (abusi di linguaggio!), ma non vale la pena essere piú precisi.

È ormai completamente chiaro che il concetto di morfismo non deve essere derivato da quello di struttura. In primo luogo i morfismi non sono univocamente definiti; in secondo luogo, non devono essere necessariamente delle applicazioni degli insiemi fondamentali, e la loro composizione non deve necessariamente essere la composizione di applicazioni. L'esempio classico, che fra

l'altro non si allontana troppo dal modello originale, è la categoria degli spazi topologici in cui si assumono come morfismi le classi di equivalenza omotopica di applicazioni continue. Un altro esempio è la categoria degli insiemi (senza ulteriore struttura) quando si prendano come morfismi le corrispondenze, cioè arbitrarie relazioni binarie. In terzo luogo, dal punto di vista categoriale, si può ritenere che la definizione della struttura sia soltanto una costruzione intermedia; quello che in realtà interessa è la categoria delle strutture di un dato tipo, ed allora è possibile che risulti piú comodo assegnare la struttura come oggetto secondario, considerando principale una certa classe importante di morfismi. Ad esempio, è naturale introdurre la struttura di topologia di Zariski sugli oggetti della categoria delle varietà algebriche, su un campo algebricamente chiuso k, come la topologia meno fine in cui k è topologizzato coi complementi degli insiemi finiti, e tutte le funzioni razionali sulle varietà sono continue al di fuori dell'insieme dei propri poli. In quarto luogo, ancora dal punto di vista categoriale, un concetto come quello di equivalenza di diversi tipi di struttura (ad esempio la nozione di gruppo assegnata con diversi sistemi di assiomi) si definisce in maniera naturale per mezzo di un funtore, che sia un'equivalenza delle corrispondenti categorie.

Esempi di questo tipo si potrebbero aumentare considerevolmente. Non si vuole tuttavia sostenere che si tratti di argomenti a favore di un totale rifiuto del concetto di struttura e di una sua sostituzione con quello di categoria. Negli articoli «Applicazioni» e «Dualità» di questa stessa *Enciclopedia* ci si è sforzati di mostrare che i concetti struttura/categoria hanno dei caratteri di complementarità, e che attualmente il miglior approccio è quello di una considerazione parallela dei due concetti, insieme alle loro complesse correlazioni e a parziali riduzioni, in casi particolari, dell'uno all'altro. In particolare, alcune classi di strutture generatrici, come quelle di gruppo, di insieme ordinato e di spazio topologico, insieme ai loro morfismi naturali, continuano a mantenere un ruolo fondamentale anche nelle costruzioni categoriali, che a loro volta permettono di considerarne le varianti piú generalizzate.

4. *Topologie di Grothendieck e topoi.*

4.1. Il programma di Grothendieck.

In questo paragrafo viene descritto un sistema di concetti sorto originariamente in geometria algebrica. Tali concetti furono introdotti da Grothendieck allo scopo di definire degli invarianti topologici delle varietà algebriche su un campo astratto, ad esempio finito. Come ha affermato Weil alla fine degli anni '40, la presenza di questi invarianti (piú precisamente la teoria della coomologia con proprietà che permettano di dimostrare una formula per il numero di punti fissi di un'applicazione) permetterebbe di capire in profondità problemi di carattere numerico (legati in particolare al numero di soluzioni di equazioni, modulo un numero primo). Questo programma venne poi realizzato (cfr. l'articolo «Geometria e topologia» in questa stessa *Enciclopedia*, al § 6) e risultò inoltre che il significato delle costruzioni del tutto generali di Grothendieck superava i limiti del problema inizialmente posto e destava un notevole interesse, in particolare presso gli specialisti di logica matematica e dei fondamenti della matematica. Attualmente le topologie e i topoi di Grothendieck e di Lawvere possono essere considerati alternativi e complementari al linguaggio strutturale, e con estrema naturalezza permettono di includere nell'ambito della matematica generale i modelli non-standard di oggetti matematici, in primo luogo degli stessi insiemi, come pure la logica intuizionista, ecc. Allo stesso tempo vengono alla luce i legami di queste costruzioni con temi da tempo assunti negli altri campi della matematica.

Per iniziare, si espongono ora le idee topologiche che conducono alla teoria di Grothendieck.

4.2. Contesto topologico.

Sia X uno spazio topologico, ad esempio una varietà differenziabile, compatta, n-dimensionale, e si supponga dato un suo ricoprimento finito $X=\bigcup_{i=1}^{N} U_i$. È noto che l'informazione essenziale sulla «forma» globale di X può essere ottenuta da una tabella finita che indichi quali intersezioni $U_{i_0}\cap\ldots\cap U_{i_k}$, $1\leq i_k\leq N$, sono non-vuote; ciò è vero, almeno, quando tutte queste intersezioni sono bolle. Piú precisamente, si scelga un gruppo abeliano A e si associ a ogni numero intero $k\geq 0$ il gruppo delle cocatene di Čech $C^k(X,(U_i);A)$, costituito dalle famiglie $\{a_{i_0\ldots i_k}\in A \mid U_{i_0}\cap\ldots\cap U_{i_k}\neq\varnothing\}$, dove $a_{i_0\ldots i_k}$ è emisimmetrica rispetto a $i_0, \ldots, i_k$. Si costruiscano gli omomorfismi $d_k : C^k\to C^{k+1}$ per mezzo delle formule: $(d_k a)_{i_0\ldots i_{k+1}}=\Sigma(-1)^i a_{i_0\ldots \hat{i}_j\ldots i_{k+1}}$ dove $\hat{i}_j$ denota che questo indice va tolto. Allora $d_{k+1}\circ d_k=0$, ossia (C^k, d_k) costituiscono un complesso. Si costruiscano i suoi gruppi di coomologia: $H^k(X,(U_i);A)=\operatorname{Ker} d_k/\operatorname{Im} d_{k-1}$. Nell'ipotesi che tutte le intersezioni non vuote $U_{i_0}\cap\ldots\cap U_{i_k}$ siano bolle, i gruppi $H^k(X,(U_i);A)$ non dipendono dalla scelta del ricoprimento (U_i) e sono invarianti topologici della varietà X. D'altro lato, come si è già osservato, sono completamente determinati da una collezione finita di dati combinatori sul ricoprimento. Senza assumere a priori condizioni addizionali sul ricoprimento, si possono definire gli omomorfismi $C^k(X,(U_i);A)\to C^k(X,(V_j);A)$ dove (V_j) è un ricoprimento piú fine di (U_i) e dimostrare che inducono degli omomorfismi dei gruppi di coomologia e quindi costruire il limite induttivo di questi gruppi di coomologia rispetto al sistema dei ricoprimenti e dei raffinamenti. Il risultato sarà lo stesso.

4.3. Motivazioni dalla geometria algebrica.

Nel tentativo di trasferire questa costruzione su varietà algebriche definite su un campo arbitrario, ci si imbatte anzitutto nella difficoltà che la topologia di Zariski è troppo grossolana. Piú precisamente, se la varietà algebrica X è ir-

riducibile, allora ogni famiglia finita di sottoinsiemi aperti non vuoti ha intersezione non vuota. Ciò significa che i gruppi di cocatene di Čech a coefficienti in A hanno un'unica struttura, dipendente da A, ma non da X, e portano a invarianti omologici banali.

La soluzione di questa difficoltà consiste nell'introdurre un gruppo «variabile» A: le coordinate delle cocatene $a_{i_0 \dots i_k}$ vanno assunte in dipendenza da $U_{i_0} \cap \dots \cap U_{i_k}$ e per di piú in maniera interna, di modo che questa dipendenza rifletta le proprietà della varietà X. Il piú semplice modo per realizzare ciò parte dall'osservazione che a ogni insieme aperto $U \subset X$ è associato il gruppo naturale $\mathcal{O}(U)$ delle funzioni razionali su X che non hanno singolarità in U. Se $a_{i_0 \dots i_k}$ assume valore in $\mathcal{O}(U_{i_0} \cap \dots \cap U_{i_k})$, è possibile definire nuovamente i gruppi $C^k(X,(U_i);\mathcal{O})$. Nella descrizione di $d_k a$ per mezzo della formula precedente occorre tener conto del fatto che $a_{i_0 \dots i_j \dots i_k}$ è la restrizione della funzione all'insieme aperto piú piccolo $U_{i_0} \cap \dots \cap \hat{U}_{i_j} \cap \dots \cap U_{i_k}$. Gli omomorfismi di C^k e di H^k, legati al passaggio a un ricoprimento piú fine, si possono definire come prima e quindi è possibile introdurre i gruppi di coomologia di Čech come limite rispetto al sistema di ricoprimenti e raffinamenti. I gruppi che si ottengono $H^k(X,\mathcal{O})$, in generale, sono già non banali e riflettono le caratteristiche omologiche essenziali di X.

Questa costruzione è suscettibile di un'ulteriore generalizzazione. Prima di tutto non c'è motivo di limitarsi a considerare le funzioni razionali su X, o generalizzazioni di questo concetto a oggetti geometrici generali quali schemi, spazi analitici o varietà differenziabili, in cui l'analogo di $\mathcal{O}$ viene detto fascio strutturale. Un conveniente contesto è fornito dalla definizione (che per comodità si darà in due tempi) di fascio di gruppi (di anelli, di moduli, di insiemi, ecc.) su uno spazio topologico:

DEFINIZIONE 1. *Si dice prefascio F sullo spazio X a valori nella categoria C (degli anelli, dei moduli, degli insiemi, ecc.) un funtore controvariante dalla categoria degli insiemi aperti di X nella categoria C. Un morfismo di insiemi aperti è un'inclusione $U \subset V$. Il corrispondente morfismo $F(V) \to F(U)$ in C si dice restrizione (da V a U).*

DEFINIZIONE 2. *Il prefascio F si dice fascio se sono verificate alcune condizioni di localizzazione: per un insieme aperto U e un suo ricoprimento aperto $U = \cup U_i$, l'oggetto $F(U)$ è univocamente determinato dalle proprie restrizioni $F(U_i)$ e si può ricostruire da una collezione $F_i \in F(U_i)$ se le restrizioni di F_i e F_j su $U_i \cap U_j$ coincidono, per ogni coppia i, j.*

In questo contesto è possibile ancora definire le cocatene e la coomologia di X a coefficienti in un fascio arbitrario F se la categoria C dei valori di questo fascio è una categoria abeliana, cioè se possiede le proprietà formali della categoria dei gruppi abeliani.

4.4. Lo slittamento creativo proposto da Grothendieck.

Questa costruzione tuttavia non risolve ancora il problema della formulazione di una teoria della coomologia adeguata alla risoluzione dei problemi aritmetici sollevati da Weil. Uno dei motivi consiste nel fatto che, in generale, conducono a coomologie interessanti i fasci di moduli sul fascio strutturale $\mathcal{O}$ di una varietà algebrica. Se questa varietà è definita su un campo di caratteristica finita, come avviene nel contesto dei problemi di Weil, allora si ottengono gruppi di coomologia della stessa caratteristica, mentre per sfruttare appieno le idee di Lefschetz è necessario avere una coomologia di caratteristica zero. Grothendieck propose una via d'uscita a questa situazione, nella quale si sottoponeva a ulteriore cambiamento lo stesso concetto di topologia. Nella variante originaria, la piú importante per le applicazioni, al posto degli insiemi aperti U di qualche spazio X si considerano dei morfismi $U \to X$ dotati di una condizione piú debole di quella di essere immersioni aperte. Si espongono ora le principali definizioni di Grothendieck nella loro forma definitiva. Per brevità si evita di menzionare l'universo in cui occorre lavorare.

4.5. Famiglie coprenti e crivelli.

Sia E una categoria. Si introduce il concetto di «topologia su E» generalizzando i dati necessari per la costruzione della coomologia di Čech che si hanno quando E è la categoria degli insiemi aperti di uno spazio topologico. Questo può essere fatto in due modi, il primo dei quali è piú vicino alla situazione classica.

Famiglie coprenti. Si dirà che su E è data una topologia se per ogni oggetto S della categoria E è definita una classe di famiglie di morfismi $(\tau_i : R_i \to S, i \in I)$, che saranno dette famiglie coprenti l'oggetto S. Ogni simile classe deve soddisfare le seguenti condizioni:

a) se $(\tau_i : R_i \to S, i \in I)$ è una famiglia coprente, per ogni morfismo $f : S' \to S$ esistono i prodotti fibrati $R'_i = R_i \times S'$ e la famiglia $(\tau'_i : R'_i \to S', i \in I)$ è coprente. Inoltre $(\mathrm{id} : S \to S)$ è una famiglia coprente;

b) se $(\tau_i : R_i \to S, i \in I)$ è una famiglia coprente e per ogni $i \in I$ è data una famiglia coprente $(\tau_{ij} : R_{ij} \to R_i, j \in I_j)$, allora $(\tau_i \circ \tau_{ij} : R_{ij} \to R, i \in I, j \in I_{ij})$ è una famiglia coprente.

Nel caso classico, come si è detto, gli oggetti di E sono insiemi aperti di uno spazio topologico, e le famiglie coprenti S sono tutti i ricoprimenti aperti di S. Una categoria con una topologia si dice *sito*.

Se è data una categoria C (nelle applicazioni si tratta della categoria degli insiemi, dei gruppi abeliani, dei moduli, ecc.), si chiama prefascio sul sito E a valori in C ogni funtore controvariante di E in C. Il prefascio F si dice fascio se sono verificate le due condizioni ulteriori:

c) per ciascuna famiglia coprente $(\tau_i : R_i \to S, i \in I)$, l'applicazione indotta $\tau : F(S) \rightrightarrows \prod_{i \in I} F(R_i)$ è un monomorfismo;

d) l'applicazione τ identifica $F(S)$ con l'ugualizzatore della coppia di applicazioni $\prod_{i \in I} F(R_i) \rightrightarrows \prod_{(i,j) \in I \times I} F(R_i \times_S R_j)$.
(Questa coppia di applicazioni corrisponde alle proiezioni di $R_i \times_S R_j$ rispettivamente sul primo e sul secondo fattore. Se $M \overset{f}{\underset{g}{\rightrightarrows}} N$ sono due applicazioni insiemistiche, si chiama loro ugualizzatore l'insieme $\{m \mid f(m) = g(m)\}$. Se si tratta di applicazioni in una categoria astratta, si chiama loro ugualizzatore un oggetto che presenta un corrispondente «funtore dei punti»: cfr. anche l'articolo «Trasformazioni naturali / categorie» in questa stessa *Enciclopedia*).

Una variante della definizione è legata al concetto di crivello.

Crivelli. Una sottocategoria E' della categoria E si dice crivello se E' contiene ogni oggetto di E per il quale esiste un morfismo verso qualche oggetto di E'. Si denoti con $\varphi(E)$ la classe di tutti i crivelli della categoria E. Ogni funtore $p : E \to F$ induce un'applicazione $\varphi(p) : \varphi(F) \to \varphi(E)$. Si può assegnare una topologia su E indicando, per ogni oggetto S della categoria E, un insieme $\mathcal{J}(S)$ di crivelli nella categoria E_S degli oggetti di E sopra S, con le proprietà:

a) per ogni morfismo $T \to S$ il rilevamento a T di ogni crivello di $\mathcal{J}(S)$ appartiene a $\mathcal{J}(T)$;
b) ogni crivello di oggetti su S, contenente qualche crivello di $\mathcal{J}(S)$, appartiene già a $\mathcal{J}(S)$;
c) ogni crivello C di oggetti sopra S, per il quale esista un crivello $R \in \mathcal{J}(S)$ tale che il rilevamento di C a ogni oggetto X di R appartenga a $\mathcal{J}(X)$, appartiene già a $\mathcal{J}(S)$.

Il passaggio dalle famiglie coprenti ai crivelli non è complicato. Alla famiglia coprente $(\tau_i : R_i \to S, i \in I)$ si associa il crivello di E_S costituito dagli oggetti $f : X \to S$ per i quali esiste un $i \in I$ tale che $\mathrm{Hom}_S(X, R_i) \neq \varnothing$.

4.6. Topoi di Grothendieck.

Sia E un sito, vale a dire una categoria con una topologia. I fasci di insiemi su E costituiscono allora una categoria che è detta topos. Viene anche detta topos (di Grothendieck) ogni categoria equivalente a una costruita in tal modo. Proprio questo concetto, e le sue modificazioni, assunsero un'importanza particolare, oltre i limiti delle esigenze puramente algebrico-geometriche, connesse soprattutto con lo studio di diverse topologie sugli schemi e delle relative teorie della coomologia, e condusse alla dimostrazione delle congetture di Weil. Il fatto è che i topoi possiedono numerose proprietà essenziali della categoria degli insiemi, «relativizzate» a una base che sia un sito.

L'esempio tipico di topos è la categoria dei fasci di insiemi sopra uno spazio topologico X. Quando X consiste di un solo punto, questo topos è equivalente alla categoria degli insiemi: ogni fascio F è rappresentato da un insieme $F(X)$. Nel caso generale ci si può rappresentare un fascio F come un «insieme variabile», i cui elementi dipendono da un sottoinsieme aperto $U \subset X$. Molto vicini ai topoi di Grothendieck sono anche i modelli non-standard a valori booleani della teoria degli insiemi, costruiti nella teoria di Cohen-Scott-Solovay (cfr. ancora l'articolo «Insieme»), per mezzo dei quali si stabilisce, ad esempio, l'indimostrabilità dell'ipotesi del continuo. Si passa ora ad esporre i concetti logici che hanno condotto a una variante dei topoi.

4.7. Topoi di Lawvere e Tierney.

Si chiamerà qui topos un oggetto assegnato mediante altri assiomi che imitano direttamente le principali costruzioni insiemistiche. Questo concetto risulta piú vicino alle necessità della logica e, nello stesso tempo, è piú ampio di quello di topos di Grothendieck.

Precisamente, si chiamerà topos una categoria T nella quale sono definiti i prodotti diretti finiti e in cui è dato un oggetto terminale 1. Oltre a ciò, a ogni oggetto X di T si associano un oggetto $P(X)$ e un monomorfismo $\varepsilon_X : \Sigma_X \to X \times P(X)$. Se T è la categoria degli insiemi, allora $P(X)$ è l'insieme delle parti di X e ε_X è l'immersione del grafico della relazione di appartenenza ε. L'assiomatizzazione delle proprietà di questo grafico conduce alla seguente postulazione che rappresenta il principale assioma dei topoi di Lawvere e Tierney:

ASSIOMA. *Per ogni coppia di oggetti X, Y e per ogni monomorfismo $i : S \to X \times Y$ esiste un unico morfismo $u : Y \to P(X)$ per il quale si ha un quadrato cartesiano*

$$\begin{array}{ccc} S & \xrightarrow{i} & X \times Y \\ {\scriptstyle j}\downarrow & & \downarrow{\scriptstyle id_X \times u} \\ \Sigma_X & \xrightarrow{\varepsilon_X} & X \times P(X). \end{array}$$

Viceversa, per ogni morfismo $u : Y \to P(X)$ esiste un unico monomorfismo i e il quadrato precedente è cartesiano.

Applicando questo assioma al caso $Y = 1$ si ottiene un morfismo $u : 1 \to P(X)$ in corrispondenza a ogni monomorfismo $i : S \to X$. Questo morfismo si dice sottoggetto di X di immagine i. In particolare l'immagine dell'applicazione identica $\mathrm{id}_X : X \to X$ è un morfismo $1 \to P(X)$ che si denota $\ulcorner X \urcorner$.

Si ponga $\Omega = P(1)$. Nella categoria degli insiemi questo è $\{\varnothing, \{\varnothing\}\}$, la coppia dei valori «falso, vero». Se nell'assioma di topos si scambiano di posto X e Y, si ha la possibilità di associare a ogni sottoggetto Y dell'oggetto X il morfismo $\varphi_Y : X \to \Omega$, il quale si dice morfismo caratteristico di Y. Viceversa ogni morfismo $\varphi : X \to \Omega$ è il morfismo caratteristico di qualche sottoggetto di X che è comodo denotare con $\{x \in X \mid \varphi(x)\}$ (nel topos degli insiemi questo è il sottoinsieme $x \in X$ per cui è vera la proprietà φ). Il morfismo caratteristico del-

l'immagine del monomorfismo diagonale $\Delta_X : X \to X \times X$ si denota con $=_X$ e il morfismo caratteristico dell'immagine di $\varepsilon_X : \Sigma_X \to X \times P(X)$ si denota con $\in_X$. Se sono dati due morfismi $f,g : X \to Y$, la composizione $X \xrightarrow{(f,g)} Y \times Y \xrightarrow{=_Y} \Omega$ è il morfismo caratteristico di qualche sottoggetto di Y che corrisponde al nucleo della coppia (f, g) e che si può chiamare ugualizzatore di questa coppia.

Dall'assioma, inoltre, si deduce che P è un funtore controvariante, come nel topos degli insiemi. Utilizzando questa proprietà e le costruzioni precedenti, si può trasformare ogni topos in un modello del linguaggio L_1 Set nel quale saranno verificati gli assiomi logici intuizionisti e una parte di quelli di teoria degli insiemi.

Per stabilire un collegamento fra i topoi di Lawvere e Tierney con quelli di Grothendieck, si introducono ora ulteriori concetti. Si indichi con $V : 1 \to \Omega$ il morfismo caratteristico del sottoggetto $\ulcorner 1 \urcorner$ dell'oggetto terminale 1, e con $\Lambda : \Omega \times \Omega \to \Omega$ il morfismo caratteristico dell'ugualizzatore dei morfismi $\mathrm{id}_\Omega \times V$, $V \times \mathrm{id}_\Omega : \Omega \times \Omega \to \Omega$. Si chiama operatore modale in T un morfismo $j : \Omega \to \Omega$ che soddisfi le seguenti condizioni: $jV = V$, $j^2 = j$, $j(x \wedge y) = jx \wedge jy$, dove x e y sono le due proiezioni $\Omega \times \Omega \to \Omega$. (Questa definizione è motivata da quelle analoghe nelle teorie delle algebre di Boole e di Heyting, relative rispettivamente ai modelli della logica usuale e di quella intuizionistica). Se è fissato un tale operatore modale j, si denoti con $\mathcal{J}$ l'ugualizzatore della coppia (j, V) e con Ω_j l'ugualizzatore della coppia (j, id_Ω). Si dice che un monomorfismo $u : X \to Y$ è chiuso (rispettivamente, denso) se esiste un quadrato cartesiano

$$\begin{array}{ccc} X & \xrightarrow{u} & Y \\ \downarrow & & \downarrow \\ \Omega_j & \longrightarrow & \Omega \end{array} \qquad \text{rispettivamente} \qquad \begin{array}{ccc} X & \xrightarrow{u} & Y \\ \downarrow & & \downarrow \\ \mathcal{J} & \longrightarrow & \Omega \end{array}$$

Un sottoggetto dell'oggetto X si dice chiuso (rispettivamente, denso) se è l'immagine di un monomorfismo chiuso (rispettivamente, denso).

Rispetto all'operatore modale j è possibile costruire la sottocategoria piena T_j del topos T costituita dagli «oggetti fascio» F che soddisfano la condizione: per ogni monomorfismo denso $u : X \to Y$ e per ogni morfismo $v : X \to F$ esiste un unico morfismo $x : Y \to F$ per il quale $v = ux$.

La categoria T_j a sua volta è un topos.

Si supponga che il topos T consista dei funtori controvarianti della categoria C nella categoria S. L'oggetto Ω associa a ogni oggetto U di C l'insieme dei crivelli in C_U ed è possibile dimostrare che è equivalente assegnare un operatore modale o una topologia nella categoria C. Il topos T_j sarà allora precisamente il topos dei fasci sul sito corrispondente C.

Se C è la categoria degli insiemi aperti di uno spazio topologico, allora il prefascio Ω associa a ogni sottoinsieme aperto U l'insieme di tutte le classi F di sottoinsiemi aperti di U, chiuse rispetto al passaggio a sottoinsiemi aperti. Si definisce il morfismo $j : \Omega \to \Omega$ in questo modo: se F è una tale classe di sottoinsiemi aperti di U, allora $j(F)$ è l'insieme dei sottoinsiemi aperti di U che contengono nella loro unione gli elementi di F. In questo esempio concreto T_j coincide con la categoria dei fasci sullo spazio topologico originario.

5. *Strutture e fisica.*

5.1. Riferimenti alla fisica teorica. Relatività generale, campo unificato e teorie quantistiche.

Nell'esposizione precedente ci si è interessati della realizzazione moderna delle idee strutturaliste in matematica, compresi i dettagli di questa realizzazione e le particolarità della sua nascita storica. Tornando a un contesto piú ampio, si tenterà di valutare il significato delle strutture matematiche per la conoscenza delle scienze naturali. Per non essere generici, si comincerà considerando brevemente le attuali relazioni fra matematica e fisica teorica.

La matematica secondo Bourbaki, la matematica delle strutture e delle categorie ha un aspetto inconsueto per la fisica teorica, e i contatti fra la nuova matematica e la nuova fisica, molto indebolitisi a partire dagli anni '40, soltanto ora cominciano a ristabilirsi. Una premessa di ciò è la rinascita del programma di Einstein relativo alla geometrizzazione della fisica a un livello, essenzialmente nuovo, su cui questo programma è considerato come un passo verso la costruzione dei fondamenti di una teoria quantistica unitaria del campo. È di Einstein il fondamentale merito di aver scoperto la struttura con cui si descrive lo spazio-tempo: una varietà quadridimensionale liscia con metrica di segnatura $(+---)$, e l'interpretazione delle proprietà della curvatura di questa metrica in termini di interazioni gravitazionali (cfr. l'articolo «Relatività» in questa *Enciclopedia*).

Parallelamente avveniva nella meccanica quantistica la scoperta delle strutture con cui si descrivono la materia a livello microscopico e le sue interazioni coi dispositivi macroscopici usati per l'osservazione. Gli sforzi congiunti di Bohr, Heisenberg, Schrödinger, Born, Dirac (per ricordare soltanto i maggiori) permisero anche di identificare queste strutture con i principali oggetti dell'analisi funzionale negli spazi di Hilbert: operatori lineari autoaggiunti e loro proprietà spettrali. In seguito, a questi si aggiunsero i metodi gruppali per la costruzione dei principali spazi e operatori, in sostituzione o a complemento del tradizionale linguaggio delle equazioni differenziali; il maggior contributo, in questo caso, fu quello di Wigner (cfr. anche l'articolo «Quanti» in questa stessa *Enciclopedia*).

La teoria fisica fondamentale che attende di essere elaborata è indubbiamente la teoria dei campi quantistici. Dal punto di vista matematico una parte rilevante dei principî di base per la sua elaborazione può essere formulata come una serie di «ricette», una parte delle quali si riferisce alla scelta di strutture della geometria differenziale (campi classici, oppure campi precedenti a una quantizzazione secondaria) e la parte rimanente si riferisce alla quantizzazione di queste strutture. Si descriverà qui piú in dettaglio la prima parte.

5.2. Strutture di geometria differenziale e loro motivazione fisica.

Nei suoi tratti piú generali, ogni aspetto della materia o di un campo si descrive con una fibrazione sullo spazio-tempo M, nel senso che una sezione di

questo fibrato si interpreta come storia virtuale dell'aspetto della materia/campo. Un metodo piú consueto per descrivere la storia di un punto materiale per mezzo della sua curva d'esistenza in M può essere ricondotto al precedente, convenendo che la curva d'esistenza del punto corrisponda a una sezione di tipo δ del fibrato banale su M avente supporto su questa curva. Tuttavia le esigenze della meccanica quantistica non permettono di limitarsi a considerare soltanto curve di esistenza (una rigida localizzazione è incompatibile col principio di indeterminazione e già una particella senza spin, con un impulso fissato, corrisponde a un'onda armonica che si svolge in tutto lo spazio), cioè sezioni del fibrato banale sullo spazio di Minkowski.

La scelta del fibrato corrispondente al dato aspetto della materia/campo viene condotta con un alto grado di affidabilità per le interazioni gravitazionali e elettromagnetiche, con un grado minore, trattandosi di un settore di ricerca, per le interazioni forti e deboli. Occorre tuttavia ricordare che, in una futura teoria unificata, anche quelle scelte che appaiono attualmente solide possono essere riviste.

La contrapposizione materia/campo ha una lunga storia. Il ruolo autonomo del campo «nel vuoto», come qualcosa in grado di influire sul movimento della materia se questa capita nel raggio d'azione del campo, fu chiaro dopo la conoscenza delle teorie di Maxwell e di Einstein. Tuttavia la teoria quantistica forzò la descrizione anche della materia (ad esempio degli elettroni) in termini di campi (per mezzo della funzione ψ). Questo, insieme alla scoperta delle proprietà corpuscolari del campo elettromagnetico, condusse quasi alla scomparsa della differenza fra «campo» e «materia». Ma nella rappresentazione contemporanea, almeno al livello dei modelli classici dei campi quantistici, tale differenza si ripresenta. Dal punto di vista fisico i quanti dei campi si suddividono in particelle fondamentali e particelle portatrici di interazioni. Rientrano fra le prime i leptoni e i quark, fra le seconde i fotoni e i gluoni. Nella descrizione geometrico-differenziale, il campo di una particella fondamentale è una sezione del fibrato corrispondente sullo spazio-tempo, mentre il campo che porta un'interazione è una connessione su questo fibrato.

Nel linguaggio della fisica, le connessioni si chiamano campi di gauge. Il piú semplice esempio di campo di gauge è il campo elettromagnetico; su di esso si vedono bene le particolarità di tali campi. Il campo elettromagnetico può essere assegnato da un vettore-potenziale, cioè da una 1-forma differenziale $A_\mu dx^\mu$, dove (x^μ) sono coordinate locali nello spazio-tempo. Tuttavia, in accordo con la trattazione classica, sono osservabili soltanto i coefficienti della 2-forma $F = dA = \Sigma F_{\mu\nu}\, dx^\mu \wedge dx^\nu$: le componenti del tensore del campo elettromagnetico. Se si identifica il campo elettromagnetico col vettore (A_μ), allora occorre considerare che le variabili dinamiche (A_μ) contengano un'informazione ridondante, non fisica: l'applicazione di gauge $A'_\mu = A_\mu + \partial_\mu \chi$ conduce a un campo fisicamente indistinguibile. Se poi si identifica il campo con il tensore $F_{\mu\nu}$, allora occorre considerare che valga l'equazione $d(dA) = 0$, cioè la condizione di chiusura della forma F.

Di fatto occorre considerare che il campo elettromagnetico sia una connessione di qualunque fibrato della materia dotata di carica. L'idea che il campo elettromagnetico fosse una connessione fu espressa nel 1918 da Weyl, ma a quel tempo si consideravano soltanto fibrati tensoriali sullo spazio-tempo e lo schema fisico concreto di Weyl non risultò valido. La sua prima realizzazione corretta si compí con la scoperta di Dirac secondo cui il campo elettrone-positrone (relativistico, ma non ancora quantistico secondario) è una sezione del fibrato spinoriale. In questo fibrato, il campo elettromagnetico induce effettivamente una connessione, la cui forma F è la forma della curvatura. Quest'interpretazione di F fu dimostrata sperimentalmente verso il 1960 mediante l'osservazione dell'effetto Aaron-Bom, che prevede l'interferenza di due fasci coerenti di elettroni intorno a una regione di spazio che racchiude un campo magnetico. Questi fasci (nell'approssimazione quasi-classica) passano attraverso una regione con campo nullo; l'interferenza è il risultato della differenza di fasi quantistiche, la quale si descrive come la curvatura del campo elettromagnetico nella regione tra i fasci.

5.3. Strutture e particelle «elementari».

Un matematico che abbia osservato lo sviluppo della fisica delle particelle elementari negli ultimi anni può riassumerne lo stato in questa forma: *a*) il tipo di struttura che regola il micromondo fisico può essere considerato relativamente stabile: si tratta del G-fibrato principale sullo spazio-tempo (G è il gruppo delle simmetrie) e dei fibrati vettoriali ad esso associati; *b*) esiste una serie di ricette pratiche per la scelta del gruppo delle simmetrie e della lagrangiana, a partire da dati sperimentali; *c*) si hanno fondate ipotesi, sebbene non definitive, per la scelta di G per le interazioni forti («cromodinamica quantistica») e per le interazioni elettriche deboli (modello di Weinberg-Salam); *d*) esistono delle ricette pratiche di passaggio da campi classici fittizi al computo degli effetti quantistici; e in queste ricette «c'è del metodo», come nella follia di Amleto. Tutto ciò è incerto e non definitivo; l'assenza di una teoria quantistica della gravitazione o, meglio, dello stesso spazio-tempo, lancia una sfida alle ricette pratiche. Gli schemi di supersimmetria o dei quanti spazio-temporali (secondo Wheeler e Hawking) si distinguono su una fisica piú «realistica», in primo luogo per la fiducia nella matematica, cioè nell'idea di Einstein per cui la compatibilità interna dei principî fondamentali conduce a una comprensione della natura piú vera rispetto a una collezione caotica di fatti. Non è dato sapere quando questa fiducia comincerà a realizzarsi.

5.4. E in futuro?

Qual è il posto delle strutture in questo quadro? Una possibile risposta consiste nel fatto che la concezione di struttura secondo Bourbaki ha rivelato per prima un ampio universo di «significato» della matematica, in grado di mediare tra il formalismo e i «significati» piú particolari delle costruzioni matematiche nelle scienze naturali, proposti insieme con le loro interpretazioni e pertanto dipendenti dalla verità di queste interpretazioni, esterne alla matematica. Alla ri-

flessione, liberata dalla necessità di credersi realtà, occorrono altri segni conduttori per non smarrirsi nella cattiva infinità del labirinto delle deduzioni formali. Il prodigio è nel fatto che questi stessi segni indicano anche alla matematica una strada utile, oltre i suoi confini. [JU. I. M.].

Bourbaki, N.
1939 *Eléments de mathématique*, I. *Théorie des ensembles. Fascicule de résultats*, Hermann, Paris.
1939 sgg. *Eléments de mathématique*, Hermann, Paris.
1948 *L'architecture des mathématiques*, in F. Le Lionnais (a cura di), *Les grands courants de la pensée mathématique*, Cahiers du Sud, Marseille; nuova ed. Blanchard, Paris 1962, pp. 35-47.
1954 *Eléments de mathématique*, I. *Théorie des ensembles* (fasc. 17), Hermann, Paris.

Coulson, C. A.
1962 *Interatomic forces: Maxwell to Schrödinger*, in «Nature», CXCV, 4843, pp. 744-49.

Dieudonné, J.
1969 *Regard sur Bourbaki*, in «Analele Universitatii Bucureşti. Seria Ştiinţele naturii: Matematică-mecanică», XVIII, 2, pp. 13-25.
1977 *Panorama des mathématiques pures. Le choix bourbachique*, Gauthier-Villars, Paris.

Fermi, E.
[1945] *A Course in Neutron Physics (part II)*, in *Note e memorie (Collected Papers)*, II. *United States 1939-54*, Accademia Nazionale dei Lincei, Roma 1965, pp. 514-41.

Halmos, P. R.
1957 *Nicolas Bourbaki*, in «Scientific American», XCVI, 5, pp. 88-99.

Landau, E.
1930 *Grundlagen der Analysis*, Akademische Verlagsgesellschaft, Leipzig.

Lawvere, F. W.
[1973] *Continuously Variable Sets: Algebraic Geometry = Geometric Logic*, in H. E. Rose e J. C. Shepherdson (a cura di), *Proceedings of the Logic Colloquium, Bristol, July 1973*, North-Holland, Amsterdam - Elsevier, New York 1975, pp. 135-56.

Il concetto di **struttura** si formalizza, all'interno della matematica, soprattutto mediante un **insieme** i cui elementi sono dotati di proprietà tecnicamente esprimibili in un **linguaggio** della **logica** matematica (per questo si veda anche **assioma/postulato**, **formalizzazione**, **matematiche**). Una maniera alternativa è quella di assegnare la struttura matematica di un oggetto mediante la sua appartenenza a una categoria (cfr. **trasformazioni naturali / categorie**) e di collegarlo mediante opportuni, o naturali, **applicazioni** a tutte le strutture dello stesso tipo. Questo atteggiamento, sorto nella geometria algebrica (cfr. **curve e superfici**, **geometria e topologia**, **invariante**, **locale/globale**), si rivela fruttuoso anche nelle considerazioni sui fondamenti della matematica, soprattutto per quanto riguarda la logica intuizionista. Attraverso il concetto di struttura la matematica si collega poi e interagisce con altre discipline, prima fra tutte con la **fisica** teorica (cfr. anche **particella**, **quanti**, **relatività**, **spazio-tempo**) e con le scienze naturali (cfr. **scienza, natura**), in generale, in cui effettua una **mediazione** costante tra il formalismo e i concetti (cfr. **concetto**) rappresentati.

echstes Kapitel,
F. Engel'sa»,
Il capitale,

975).

e

Trasformazioni naturali / categorie

«Le buone teorie generali non ricercano il massimo di generalità, ma la giusta generalità».

[MacLane 1971, trad. it. p. 133].

1. *Sul concetto di naturalità.*

La riflessione sul carattere piú o meno naturale, spontaneo o piuttosto innaturale o artificioso di enunciati, dimostrazioni o procedimenti matematici è da sempre intimamente connessa allo sviluppo della matematica stessa. È dunque difficile trattare l'argomento in generale, ma per introdurre adeguatamente il concetto di «trasformazione naturale» può essere utile vedere come molti elementi del discorso si siano come cristallizzati intorno a una pur semplice proposizione: si tratta della quinta proposizione del primo libro degli *Elementi* di Euclide.

Eccone l'enunciato, nella versione di Frajese e Maccioni [1970, p. 85]:

PROPOSIZIONE. *Nei triangoli isosceli gli angoli alla base sono uguali fra loro, e venendo prolungati i lati uguali gli angoli sotto la base saranno [pure] uguali fra loro.*

La dimostrazione (cfr. fig. 1) consiste nell'assumere i punti F e G, sui prolungamenti di AB e di AC, con $AF = AG$, e nell'osservare in primo luogo che i triangoli AFC e AGB sono uguali, il che si ricava da quanto precede nel testo. Da questo si conclude che anche i triangoli BFC e CGB sono uguali, e dunque anche l'angolo BCF è uguale all'angolo CBG. D'altro lato già s'è visto che l'angolo ABG è uguale all'angolo ACF e, per differenza di angoli uguali, si conclude infine che gli angoli alla base sono uguali. Implicitamente già s'è visto che anche gli angoli sotto la base sono uguali.

Certo tutto ciò che si utilizza nella dimostrazione ha carattere assai elemen-

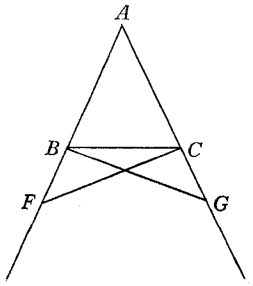


Figura 1.
L'uguaglianza degli angoli alla base di un triangolo isoscele.

tare, ma è davvero necessario passare attraverso questa catena di uguaglianze? Non può darsi una piú semplice dimostrazione, ad esempio con la considerazione della bisettrice dell'angolo BAC, come suggeriscono, in nota, Frajese e Maccioni? Per quale ragione Euclide non ha scelto una dimostrazione piú facile? Si tratta forse, come osservano Frajese e Maccioni, d'un «*pezzo forte* in favore della teoria di Zeuthen sulla dimostrazione di esistenza, presso i Greci, mediante la costruzione» [*ibid.*]?

In realtà, critiche a questa dimostrazione si trovano già nell'antichità. Proclo [*Commento al primo libro degli Elementi di Euclide*, 249, 20 - 250 sgg.] riferisce d'una dimostrazione di Pappo: «Pappo tuttavia svolge la dimostrazione in modo piú conciso, senza bisogno di alcuna aggiunta: sia ABC il triangolo isoscele e la AB sia uguale alla AC. Immaginiamo quest'unico triangolo come se fosse due e ragioniamo cosí: poiché AB è uguale ad AC, e AC è uguale ad AB, le due rette AB, AC sono uguali alle due AC, AB, e l'angolo BAC è uguale all'angolo CAB, perché sono lo stesso. Perciò le parti sono tutte uguali a tutte, la BC è uguale alla CB, il triangolo ABC è uguale al triangolo ACB, l'angolo ABC è uguale all'angolo ACB, e l'angolo ACB è uguale all'angolo ABC, perché sono sottesi dai lati uguali AB, AC; per conseguenza gli angoli alla base dei triangoli isosceli sono uguali». Il senso dell'argomentazione di Pappo può dunque riassumersi dicendo che, a suo giudizio, la dimostrazione euclidea ha un carattere artificioso, complicato, dal momento che ne esiste un'altra ben piú evidente.

Una celebre osservazione di Vico trae tutt'altre conseguenze: «Però, osservando il Vico cosí da Aristotile come da Platone usarsi assai sovente pruove mattematiche per dimostrare le cose che ragionano essi in filosofia, egli in ciò si vide difettoso a poter bene interdergli; onde volle applicarsi alla geometria e inoltrarsi fino alla quinta proposizione di Euclide [che è quella in esame]. E, riflettendo che in quella dimostrazione si conteneva insomma una congruenza di triangoli esaminata partitamente per ciascun lato ed angolo di triangolo, che si dimostra con egual distesa combaciarsi con ciascun lato ed angolo dell'altro, pruovava in se stesso cosa piú facile l'intendere quelle minute verità tutte insieme, come in un genere metafisico, di quelle particolari quantità geometriche. E a suo costo sperimentò che alle menti già dalla metafisica fatte universali non riesce agevole quello studio propio degli ingegni minuti, e lasciò di seguitarlo» [1725-28, ed. 1977 pp. 14-15].

Come ben si vede, Vico argomenta a partire da questa proposizione sull'intero carattere del dimostrare matematico. Anche un altrettanto celebre giudizio di Kant, pur di segno contrario, è dello stesso tenore: «Innanzi a colui che dimostrò i primi teoremi sul triangolo isoscele (fosse Talete o chiunque altro) si accese una gran luce, poiché comprese che non doveva seguire ciò che via via vedeva nella figura, né attenersi al semplice concetto della figura stessa, quasi dovesse apprenderne le proprietà; ma doveva produrre la figura (costruendola) secondo ciò che con i suoi concetti pensava e rappresentava in essa a priori; comprese cioè che per sapere con sicurezza qualcosa a priori, non doveva attribuire alla cosa se non ciò che risultava necessariamente da quanto, conformemente al suo concetto, egli stesso vi aveva posto» [1787, trad. it. pp. 41-42].

Qui non interessa il carattere che Kant giudica proprio del dimostrare matematico; si desidera portare l'attenzione sulla «gran luce», sulla verità che questo teorema è in grado di arrecare, indipendentemente dalla tecnica dimostrativa. In qualche modo, il giudizio di Vico è ribaltato, ma con una considerazione della matematica altrettanto globale, pur a partire da una singola proposizione. Luce od oscurità si riverberano sulla totalità, perché v'è omogeneità fra la parte e il tutto. A proposito di questo problema è interessante rilevare come la scrittrice Lalla Romano in *Una giovinezza inventata* (1979, cap. XVII) riporti un gustoso aneddoto relativo a Peano sulla sua insofferenza per l'espressione «dimostrò il triangolo isoscele». Non è inopportuno accostare quest'episodio agli esempi che si stanno qui svolgendo.

L'ultimo autore che si vuole considerare è Bergson: «Ma allorquando io traccio grossolanamente sulla sabbia la base di un triangolo, e comincio a formare i due angoli alla base, so in maniera certa e comprendo assolutamente che, se questi due angoli sono eguali, anche i lati lo saranno, poiché la figura può rigirarsi su se stessa senza che nulla sia alterato. Lo so ben prima d'aver appreso la geometria. Cosí anteriormente alla geometria dotta, vi è una geometria naturale la chiarezza e l'evidenza della quale superano quella delle altre deduzioni» [1907, ed. 1970 p. 674]. Anche Bergson, in qualche modo, trascura gli aspetti dimostrativi nel far risalire la verità a un fatto intuitivo. È però evidente che nel riproporre nella sostanza l'idea di Pappo, implicitamente giudica in favore della sua dimostrazione.

Gli esempi qui riportati, sostanzialmente uniti solo dal riferirsi a una stessa proposizione d'Euclide, mostrano una pluralità di idee e di giudizi. Ma preme osservare come molti elementi che si troveranno nelle pagine seguenti siano già impliciti: l'idea che si possa distinguere tra la verità d'un enunciato e il *modo* con cui esso si rivela per vero, la necessità d'una «coerenza» tra le singole proposizioni, d'un ugual livello dimostrativo; oppure l'indifferenza verso questi aspetti e l'enfasi sul risultato. Tutto questo, in modo implicito o esplicito, apparirà nel seguito.

2. *Trasformazioni naturali.*

Il concetto di trasformazione naturale è stato elaborato per la prima volta da Eilenberg e MacLane [1942] studiando l'isomorfismo tra un gruppo finito e il suo doppio gruppo dei caratteri. L'argomento, ripreso in forma piú organica dai due autori [1945], conduce alla prima esposizione sistematica della teoria delle categorie (si veda anche l'articolo «Funzioni» in questa stessa *Enciclopedia*, VI, p. 491). Non c'è modo migliore per ripresentare organicamente l'argomento di quello di rifarsi all'esempio addotto dai due autori.

Sia L uno spazio vettoriale di dimensione finita sul campo reale; il suo spazio duale $T(L)$ è allora lo spazio vettoriale delle applicazioni da L a $\mathbf{R}$ (si veda anche l'articolo «Dualità» in questa stessa *Enciclopedia*, V, pp. 136-37). Come

è ben noto, lo spazio $T(L)$ è a sua volta uno spazio vettoriale su **R** della stessa dimensione di L. Gli spazi L e $T(L)$ sono dunque isomorfi. Ma questo isomorfismo non può essere esplicitato se non scegliendo una base per lo spazio L; inoltre scelte differenti della base conducono a isomorfismi differenti. L'isomorfismo che collega L a $T(L)$ è dunque legato a una nostra scelta.

La situazione cambia completamente se si considera il doppio duale $T(T(L))$, che si indicherà con $T^2(L)$. In questo caso l'isomorfismo che collega L a $T^2(L)$ può essere dato direttamente, senza ricorrere alla scelta di alcuna base in L; la scelta è inoltre data in modo simultaneo per tutti gli spazi vettoriali di dimensione finita. Se infatti si considera un vettore x di L, per ogni applicazione lineare t di $T(L)$ risulta definito un numero reale calcolando t su x, cioè $t(x)$. Se ora si immagina, in $t(x)$, non x ma t variabile, si ottiene un'applicazione da $T(L)$ a **R**. Ogni elemento di L è dunque interpretabile come un'applicazione da $T(L)$ a **R**, cioè come un elemento di $T^2(L)$. In tal modo si ottiene un'applicazione $\tau(L): L \to T^2(L)$ che, come si vede facilmente, è un isomorfismo (si veda ancora «Dualità», p. 137).

In che senso questo isomorfismo è «naturale»? Intuitivamente, si può dire che, poiché la determinazione di $\tau(L)$ è data da L stesso, senza alcun intervento di scelta da parte nostra la naturalità si rivelerà tale «al variare di L».

In termini precisi, sia $\lambda: L \to L'$ un'applicazione lineare fra due spazi vettoriali di dimensione finita (il modo con il quale si descrive la variazione in questo contesto); essa induce una corrispondente applicazione fra $T^2(L)$ e $T^2(L')$. Si consideri infatti il diagramma $L \xrightarrow{\lambda} L' \xrightarrow{t'} \mathbf{R}$: esso suggerisce come per ogni applicazione $t': L' \xrightarrow{t'} \mathbf{R}$ si abbia, per composizione, un'applicazione $\lambda t': L \to \mathbf{R}$. Cioè l'operazione di comporre con λ induce un'applicazione che si dirà $T(\lambda): T(L') \to T(L)$, la quale è facilmente vista essere lineare a sua volta (T è un «funtore controvariante»; cfr. oltre, e anche l'articolo «Applicazioni» in questa stessa *Enciclopedia*, I, p. 741).

Data $T(\lambda): T(L') \to T(L)$, utilizzando il medesimo procedimento si otterrà $T^2(\lambda): T^2(L) \to T^2(L')$.

Si dispone ora di tutti gli elementi; precisamente: *a*) per ogni L, un isomorfismo $\tau(L): L \to T^2(L)$; *b*) per ogni applicazione lineare $\lambda: L \to L'$, una corrispondente applicazione lineare $T^2(\lambda): T^2(L) \to T^2(L')$.

Con questi dati, è possibile considerare il diagramma

$$(1) \qquad \begin{array}{ccc} L & \xrightarrow{\tau(L)} & T^2(L) \\ {\scriptstyle\lambda}\downarrow & & \downarrow{\scriptstyle T^2(\lambda)} \\ L' & \xrightarrow{\tau(L')} & T^2(L') \end{array}$$

ed ecco ora come si può esplicitare il concetto di naturalità: il diagramma (1) è commutativo, ossia, in termini di composizioni di applicazioni, $\lambda \cdot \tau(L') = \tau(L) \cdot T^2(\lambda)$.

La verifica della commutatività di (1) segue facilmente dall'interpretare i simboli delle applicazioni che vi compaiono. La verifica del fatto che la commutatività di (1) sia una spiegazione «adeguata» non è invece affidabile a un sem-

plice calcolo. In parte sarà trattata in questo articolo, per la piú gran parte si fonda sull'esistere concreto della teoria delle categorie.

Intanto, l'esempio stesso è suscettibile di un'indagine piú approfondita, che consenta di cogliere elementi generali, al di là di questa particolare situazione. Si può intanto isolare come un momento concettuale a sé della nostra spiegazione il passaggio da L a $T^2(L)$, evidenziandone le proprietà. Tale passaggio agisce a due livelli: al livello degli spazi vettoriali, poiché per ogni spazio L dà uno spazio $T^2(L)$; al livello delle applicazioni tra spazi vettoriali, poiché per ogni $\lambda: L \to L'$ dà $T^2(\lambda): T^2(L) \to T^2(L')$. Si può poi mostrare come le proprietà che agiscono in modo essenziale siano:

1) all'identità $\mathbf{1}_L: L \to L$, T^2 (si indica con T^2 l'intero processo) associa l'identità $\mathbf{1}_{T^2(L)}: T^2(L) \to T^2(L)$;
2) alla composizione di due applicazioni λ, λ', T^2 associa la composizione di $T^2(\lambda)$ e di $T^2(\lambda')$; cioè $T^2(\lambda\lambda') = T^2(\lambda)T^2(\lambda')$.

La 1) e la 2) definiscono T^2 come un funtore (cfr. ancora «Applicazioni», p. 741) che agisce dalla totalità degli spazi vettoriali di dimensione finita; la si indichi con **V**, in se stessa: $T^2: \mathbf{V} \to \mathbf{V}$. (Due osservazioni: non v'è possibilità d'errore nell'indicare l'agire di T^2 ancora con una freccia. Inoltre, la «totalità» degli enti di un certo tipo è naturalmente problematica anche nel contesto della teoria delle categorie. Non si affronteranno qui tali questioni, rinviando per esempio a MacLane [1971, trad. it. pp. 36-40] o a Lolli [1977] e si farà liberamente uso della terminologia *naïve*).

È opportuno anche pensare la situazione $\lambda: L \to L'$ come descritta da un funtore: naturalmente sarà il funtore identico $I: \mathbf{V} \to \mathbf{V}$ per il quale $I(L) = L$ e $I(\lambda) = \lambda$.

Si dispone ora di due funtori T^2 e I da **V** a **V** ed è possibile definire τ a partire dai $\tau(L)$ come una trasformazione naturale da I a T^2. Si scriverà $\tau: I \dashrightarrow T^2$. Il diagramma (1) diviene ora

$$(2) \qquad \begin{array}{ccc} I(L) & \xrightarrow[\tau(L)]{} & T^2(L) \\ {\scriptstyle I(\lambda)}\downarrow & & \downarrow{\scriptstyle T^2(\lambda)} \\ I(L') & \xrightarrow[\tau(L')]{} & T^2(L') \end{array}$$

che è la definizione di τ come trasformazione naturale fra i due funtori I e T^2, ossia il «dato» di tutte le applicazioni $\tau(L)$, per ogni L, con per ogni $\lambda: L \to L'$ la commutatività di (2).

Legati ancora all'esempio particolare, si sono tuttavia evidenziati due concetti generali, quello di funtore e quello di trasformazione naturale. Un nuovo concetto scaturisce dall'indagare quali proprietà in realtà siano richieste al «dominio» e al «codominio» di un funtore: il concetto di categoria.

Una definizione formale di categoria è ricorsa piú volte negli articoli di questa *Enciclopedia* (si vedano «Applicazioni», p. 736, nel contesto insiemistico, e «Funzioni», p. 491, con una definizione analoga a quella di «metacategoria» [cfr. anche MacLane 1971, trad. it. pp. 19-20]). Disponendo di essa è facile

adattare la definizione di funtore da **V** a **V** al caso generale di due categorie **A** e **B**. Allora, se F e G sono due funtori $\mathbf{A} \xrightarrow{F} \mathbf{B}$, $\mathbf{A} \xrightarrow{G} \mathbf{B}$, una trasformazione naturale $\tau : F \dashrightarrow G$ è assegnata dando per ogni oggetto A di **A** un morfismo $\tau_A : F(A) \to G(A)$ tale che per ogni morfismo $f : A \to B$ il diagramma

$$\begin{array}{ccc} F(A) & \xrightarrow{\tau_A} & G(A) \\ {\scriptstyle F(f)}\downarrow & & \downarrow{\scriptstyle G(f)} \\ F(B) & \xrightarrow{\tau_B} & G(B) \end{array}$$

sia commutativo (cfr. ancora «Dualità», pp. 155-56).

Si osserverà come in questa presentazione funtori e categorie appaiano come concetti subordinati per esprimere la naturalità. E tale era certo agli inizi la situazione teorica. Va osservato tuttavia che in seguito la situazione si è equilibrata poiché le strutture descritte come categorie hanno rivelato intrinsecamente grandi capacità euristiche. Oggi si parla senz'altro di teoria delle categorie, forse anche perché le trasformazioni naturali non sono l'unico caso ove si esplica il concetto di naturalità (si veda in seguito, ad esempio, il concetto di coppia di funtori aggiunti).

Per concludere, ecco ora alcune considerazioni. Osserva MacLane: «Una trattazione diretta delle categorie in quanto tali apparve in Eilenberg e MacLane (1945). Ora la scoperta d'idee generali come queste è soprattutto la propensione a fare un'astrazione avventata o di carattere speculativo, in questo caso sostenuta dal piacere di sottrarre parole ai filosofi: "categoria" ad Aristotele e Kant, "funtore" a Carnap (*Logische Syntax der Sprache*), e "trasformazione naturale" al linguaggio informale allora corrente» [*ibid.*, p. 46]. Il tono scanzonato non deve trarre in inganno (non si dimentichi fra l'altro che MacLane è stato allievo di Bernays): fin dall'inizio la teoria si presenta con precise istanze filosofiche. Al termine dell'introduzione del loro articolo Eilenberg e MacLane osservavano: «Il carattere invariante di una disciplina matematica può essere formulato in questi termini. Cosí nella teoria dei gruppi tutte le costruzioni basilari possono essere riguardate come definizioni di funtori covarianti o controvarianti, cosí che è possibile formulare il detto: l'argomento della teoria dei gruppi è essenzialmente lo studio di quelle costruzioni gruppali che si comportano in maniera covariante o controvariante per gli omomorfismi indotti. Piú precisamente, la teoria dei gruppi studia funtori definiti su ben specificate categorie di gruppi, con valori in altre categorie simili. Questo può essere considerato come una continuazione del Programma di Erlangen di Klein, nel senso che uno spazio geometrico con il suo gruppo di trasformazioni è generalizzato a una categoria con la sua algebra di morfismi» [1945, p. 237].

Ma nell'immediato i fatti non parvero accordarsi con le previsioni. Afferma ancora MacLane: «Nel successivo decennio 1945-55 le categorie sono state poco studiate, la teoria delle categorie era solo un linguaggio, ed eventuali ricercatori possono essere stati scoraggiati dal diffuso dubbio pragmatico di "*nonsense* astratto generale" (teoria delle categorie)» [1971, trad. it. p. 133]. In realtà la situazione pareva suggerire che le principali istanze del concetto di categoria

si riducessero a insiemi con una certa struttura e loro morfismi (gruppi, anelli, moduli, ecc.). Era poco operante l'idea di utilizzare una descrizione *intrinseca* delle categorie stesse (sebbene presente fin dall'inizio), e mancavano esempi significativi di categorie non descrivibili come in precedenza. La situazione cambierà progressivamente, verso la fine degli anni '50, con la scoperta da parte di Grothendieck (1957) «che le categorie di fasci (di gruppi abeliani) sopra uno spazio topologico erano categorie abeliane ma non categorie di moduli, e che l'algebra omologica in queste categorie era necessaria per una trattazione completa della coomologia in un fascio (Godement, 1958)» [*ibid.*, p. 256]. A questa si affiancherà, nel 1958, la scoperta da parte di Daniel Kan dei funtori aggiunti.

Per procedere in modo non banale nell'esame degli sviluppi successivi occorre ora illustrare qualche costruzione categoriale.

3. *Qualche costruzione categoriale.*

Le costruzioni che vengono qui esposte risultano molto chiare se vengono pensate con l'idea-guida che i morfismi $X \xrightarrow{x} A$, che hanno per codominio un oggetto A di una categoria **C**, debbano essere considerati come «elementi generalizzati di A». Sviluppare quest'idea, dandole concretezza, richiede un approfondimento teorico che, in parte, sarà sviluppato nelle pagine seguenti. Ci si limiterà per ora a far notare come quest'idea sia una generalizzazione dell'abituale nozione di elemento negli insiemi.

Sia **S** (dall'inglese *set*) la categoria degli insiemi [si rinvia ancora a MacLane 1971 o a Lolli 1977 per il senso con il quale si devono intendere «tutti gli insiemi», nel contesto categoriale] e si consideri in essa l'insieme formato da un solo elemento $\{*\}$. Si indica abitualmente tale insieme con 1. La caratterizzazione di 1 in termini di morfismi è data dal fatto che, per ogni insieme A, esiste esattamente un morfismo $A \to 1$, cosa di immediata verifica. Complementarmente, gli elementi di A corrispondono esattamente ai morfismi $1 \to A$, nel senso preciso che c'è una corrispondenza biunivoca fra questi morfismi e gli elementi di A.

Due generalizzazioni si presentano ora possibili: per ogni categoria **C**, ove vi sia un oggetto come 1 (oggetto terminale), è possibile definire gli elementi di un oggetto A come i morfismi $1 \to A$. Si parla in tal caso di «elementi globali» o «eterni» (Lawvere) di A: una generalizzazione utile, ma piú utile è quella che si è predescritta, ove viene generalizzato anche il dominio dei morfismi di codominio A.

Ecco ora qualche applicazione. Negli insiemi, una funzione f, con $f : A \to B$, è detta iniettiva quando, dati due elementi a, a' di A, con $a \neq a'$, si ha di conseguenza $f(a) \neq f(a')$. O anche: $f(a) = f(a')$ implica $a = a'$. Questa stessa definizione, data nel contesto categoriale, con la definizione di elemento generalizzato conduce al concetto di monomorfismo (mono).

DEFINIZIONE. *$f: A \to B$ è un monomorfismo se, per ogni coppia di morfismi $a: X \to A$, $a': X \to A$,*

$$X \underset{a'}{\overset{a}{\rightrightarrows}} A \xrightarrow[f]{} B$$

l'uguaglianza $af = a'f$ implica $a = a'$.

Nel già citato articolo «Funzioni» si è descritto (p. 492) il prodotto cartesiano cercando soprattutto di mettere in evidenza la sua proprietà universale. È facile ora rendersi conto del fatto che, con questa nuova definizione di elemento, essa consiste esattamente nel dire che il prodotto cartesiano è l'insieme delle coppie di elementi (generalizzati).

Nello stesso spirito, si veda ora l'ugualizzatore. Negli insiemi, date due funzioni f, g aventi lo stesso dominio A e lo stesso codominio B, $A \underset{g}{\overset{f}{\rightrightarrows}} B$ si può definire quel sottoinsieme S di A formato dagli elementi x di A tali che $f(x) = g(x)$. Il sottoinsieme S di A, nel contesto che qui interessa, va pensato come un monomorfismo $m: S \to A$ [per una profonda discussione sul fatto che il concetto di sottoinsieme $S \subseteq A$ è «troppo astratto» e che un concetto piú adeguato è quello di «essere sottoinsieme in modo determinato», ossia l'assegnazione di un monomorfismo $m: S \to A$ che precisi il «modo» in cui S è sottoinsieme di A si rinvia a Lawvere 1972-73. L'argomento è efficacemente ripreso in Meloni 1974-77]. Ora, ogni elemento di A che renda uguali f e g è già in S. Ciò motiva la seguente

DEFINIZIONE. *L'ugualizzatore di due morfismi $f: A \to B$ e $g: A \to B$ è un morfismo $m: S \to A$ tale che, per ogni morfismo $x: X \to A$ con $xf = xg$*

$$\begin{tikzcd} S \arrow[dr, "m"] & & \\ & A \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & B \\ X \arrow[ur, "x"'] \arrow[uu, "h"] & & \end{tikzcd}$$

vi sia esattamente un morfismo h con $hm = x$.

(Nella definizione, si preferisce richiedere l'unicità di h piuttosto che m sia un monomorfismo).

In modo analogo si potrebbero motivare ora altre costruzioni. Per non essere troppo didascalici si presenta direttamente la definizione di prodotto fibrato.

DEFINIZIONE. *Dati due morfismi $f: A \to X$ e $g: B \to X$*

$$\begin{tikzcd} & B \arrow[d, "g"] \\ A \arrow[r, "f"'] & X \end{tikzcd}$$

il prodotto fibrato dei due morfismi è dato da un oggetto Y e da una coppia di morfismi $\pi_1: Y \to A$ e $\pi_2: Y \to B$ tali che il quadrato

$$\begin{tikzcd} Y \arrow[r, "\pi_2"] \arrow[d, "\pi_1"'] & B \arrow[d, "g"] \\ A \arrow[r, "f"'] & X \end{tikzcd}$$

sia commutativo. Inoltre, se $Z \xrightarrow{h} A$ e $Z \xrightarrow{k} B$ sono altri due morfismi per i quali $hf = kg$, allora esiste un unico morfismo $\varphi: Z \to Y$

$$\begin{tikzcd} Z \arrow[dr, "\varphi"] \arrow[drr, bend left, "k"] \arrow[ddr, bend right, "h"'] & & \\ & Y \arrow[r, "\pi_2"] \arrow[d, "\pi_1"'] & B \arrow[d, "g"] \\ & A \arrow[r, "f"'] & X \end{tikzcd}$$

tale che $\varphi\pi_1 = h$ e $\varphi\pi_2 = k$.

I concetti di monomorfismo, ugualizzatore, prodotto cartesiano e prodotto fibrato, or ora presentati, costituiscono dispositivi essenziali della teoria delle categorie. Essi possono tutti essere unificati come «limiti finiti» [si veda MacLane 1971, trad. it. cap. III]. Accanto a questi si dànno i concetti «duali», ottenuti pensando ad essi nella categoria opposta (con gli stessi oggetti e le frecce in senso contrario). A titolo d'esempio:

DEFINIZIONE. *Dati i due morfismi $f: A \to B$ e $g: A \to B$ il loro cougualizzatore è dato da un morfismo $p: B \to T$ tale che, per ogni morfismo $y: B \to Y$ con $fy = gy$ vi sia esattamente un morfismo $k: T \to Y$ con $pk = y$.*

$$\begin{tikzcd} A \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & B \arrow[r, "p"] \arrow[dr, "y"'] & T \arrow[d, "k"] \\ & & Y \end{tikzcd}$$

In modo analogo si hanno i concetti di epimorfismo (epi), duale di quello di monomorfismo, di coprodotto, duale di quello di prodotto e di *coprodotto fibrato*, duale di quello di prodotto fibrato. In seguito si farà liberamente uso di tutti questi concetti.

4. *Funtori aggiunti.*

Il concetto di coppia di funtori aggiunti, isolato per la prima volta da Kan [1958], è certo fondamentale nella teoria delle categorie e corrisponde a una pluralità di situazioni molto diversificata e articolata (e proprio la sua capacità unificante dà la sua gran carica concettuale).

Esso non è quindi ricavabile come generalizzazione da un esempio principale, come è stato fatto in precedenza. È piú opportuno quindi premettere la definizione astratta e mostrare poi con successivi esempi la ricchezza di contenuti.

DEFINIZIONE. *Siano $\mathbf{A}$ e $\mathbf{B}$ due categorie e F, G due funtori con $F: \mathbf{A} \to \mathbf{B}$ e $G: \mathbf{B} \to \mathbf{A}$. Si dirà che F è l'aggiunto sinistro di G (e si scriverà $F \dashv G$) se, per ogni coppia di oggetti A in $\mathbf{A}$ e B in $\mathbf{B}$, è dato un isomorfismo $\varphi_{A,B}$, naturale in A e B tra i due insiemi di morfismi $\mathbf{B}(F(A), B)$ e $\mathbf{A}(A, G(B))$,*

$$\varphi_{A,B}: \mathbf{B}(F(A), B) \simeq \mathbf{A}(A, G(B)).$$

(Con la notazione $\mathbf{C}(X,Y)$ si indica, dati X e Y oggetti di una categoria $\mathbf{C}$, l'insieme dei morfismi da X a Y. La definizione di funtori aggiunti si trova anche nel già citato articolo «Dualità», funzionale tuttavia a un discorso diverso).

Di grande aiuto, forse per la persuasività dell'immagine, è raffigurare la situazione di aggiunzione con lo schema

$$\frac{A \xrightarrow{\varphi(f)} G(B)}{F(A) \xrightarrow{f} B}$$

(si è semplificato indicando $\varphi_{A,B}$ con φ). La naturalità rispetto agli oggetti A e B può essere descritta in termini di trasformazioni naturali introducendo i funtori adeguati dalla categoria prodotto di $\mathbf{A}^{\mathrm{op}}$ e di $\mathbf{B}$ in $\mathbf{S}$. Ma se ne può dare una descrizione molto semplice utilizzando la figura seguente:

$$\frac{A' \xrightarrow{h} A \xrightarrow{\varphi(f)} G(B) \xrightarrow{G(k)} G(B')}{F(A') \xrightarrow[F(h)]{} F(A) \xrightarrow[f]{} B \xrightarrow[k]{} B'}.$$

La naturalità si può ora esprimere con $\varphi(F(h)f) = h\varphi(f)$ e con $\varphi(f \cdot k) = \varphi(f)\,G(k)$. Ecco ora alcuni esempi.

Esempio 1.

Sia Vet la categoria degli spazi vettoriali su un determinato campo K, e $U: \mathrm{Vet} \to \mathbf{S}$ il funtore «dimenticante», che considera cioè per ogni spazio vettoriale V i vettori di V semplicemente come un insieme (dimenticando la struttura). Sia $U(V)$ tale insieme. Sia B il funtore che a ogni insieme X associa lo spazio vettoriale $B(X)$ con base X su K. Allora, per ogni insieme X e ogni spazio V si ha

$$\frac{X \to U(V)}{B(X) \to V}$$

cioè $B \dashv U$, poiché un morfismo tra spazi vettoriali è individuato esattamente assegnando le immagini dei vettori della base. La naturalità, come si vede, significa che la corrispondenza è definita *in modo uniforme* per X e V.

Esempio 2.

Si consideri la categoria degli insiemi $\mathbf{S}$. Dati due oggetti A e B è possibile formare il loro prodotto cartesiano $A \times B$; dato ancora un oggetto C, è possibile considerare anche B^C, la totalità delle funzioni da C a B. Si ha ora

$$\frac{A \to C^B}{A \times B \to C}$$

a causa del procedimento familiare con il quale una funzione di due variabili si può ridurre a una sola variabile cambiando in modo opportuno il codominio. Questo significa che il funtore $(\ ) \times B$, definito dall'effettuare il prodotto cartesiano di un insieme per B e in modo conseguente sui morfismi, è aggiunto sinistro del funtore di altrettanto ovvia descrizione $(\ )^B : (\ ) \times B \dashv (\ )^B$.

Esempio 3.

Siano $\mathbf{A}$ e $\mathbf{B}$ due preordini (cioè categorie tali che per ogni coppia di oggetti vi sia al piú un morfismo tra questi). Allora la situazione

$$\frac{A \to G(B)}{F(A) \to B}$$

significa $A \leq G(B)$ se e solo se $F(A) \leq B$. Nel caso $\mathbf{B} = \mathbf{C}^{\mathrm{op}}$ si ha una connessione di Galois.

Esempio 4.

È questo l'esempio piú importante (almeno per chi scrive). Si tratta della descrizione dei quantificatori come funtori aggiunti [Lawvere 1969].

Si consideri una situazione matematica ove intervenga uno dei quantificatori $\exists$, $\forall$. Si immagini per esempio di passare dalla formula $xy = 1$ alla formula $\exists x(xy = 1)$.

Passare dalla prima alla seconda formula è anche un fatto «linguistico»: una formula diviene un'altra formula. Si possono dare delle regole che consentano di dire che la seconda esprime in modo adeguato il concetto «esiste un x tale che...» (cfr. l'articolo «Logica» in questa stessa *Enciclopedia*, VIII, pp. 506 sgg.). Ma si utilizza qui un punto di vista estensionale. x e y saranno variabili intese a indicare gli elementi di un certo insieme A, per esempio un anello. Allora con $xy = 1$ si individua quel sottoinsieme di $A \times A$ dato dalle coppie (x,y), le quali sono formate appunto dagli elementi il cui prodotto è l'unità

$$\begin{array}{c} S \\ \downarrow \\ A \times A \end{array} \qquad S = \{(x,y) \mid xy = 1\}.$$

Si può ora considerare la funzione «seconda proiezione» da $A \times A$ ad A definita con $\pi_2(x,y) = y$.

È possibile infine considerare il sottoinsieme S' di A dato da quegli y che provengono mediante π_2 dagli elementi di S

$$(3) \qquad \begin{array}{ccc} S & & S' \\ \downarrow & & \downarrow \\ A \times A & \xrightarrow{\pi_2} & A \end{array}$$

Sono, naturalmente, gli elementi di A per i quali esiste un x con $xy = 1$.

Tutta l'azione del quantificatore può allora essere descritta come l'associare a $S \xrightarrow{s} A \times A$ (si evidenzia ancora il modo con cui S è contenuto in $A \times A$) un $S' \xrightarrow{s'} A$. Come è definito s'? Esso non è che l'immagine della composizione $s\pi_2$. Per mettere in evidenza questo fatto si scriverà $s' = \exists_{\pi_2}(s)$. Ma il modo stesso della definizione suggerisce una generalizzazione. Si consideri

$$\begin{array}{c} S \\ {\scriptstyle s}\downarrow \\ X \xrightarrow[f]{} Y \end{array}$$

ove X sostituisce $(A \times A)$ e Y (A). Non è altrettanto facilmente definibile $s' : S' \to Y$ come l'immagine di sf? Questa definizione è tuttavia suscettibile di essere riformulata in maniera ancora piú profonda. Il morfismo $f : X \to Y$ definisce per ogni morfismo $h : R \to Y$ attraverso il prodotto fibrato

$$\begin{array}{ccc} Q & \longrightarrow & R \\ {\scriptstyle f^*(h)}\downarrow & & \downarrow{\scriptstyle h} \\ X & \xrightarrow[f]{} & Y \end{array}$$

un nuovo morfismo $f^*(h) : Q \to X$. Nel caso degli insiemi è evidentemente l'immagine inversa di $h : R \to Y$, ma, si osserverà, esso è definibile in ogni categoria ove esistano i prodotti fibrati. Negli insiemi, poi, indicando l'immagine di sf con $\exists_f(S)$ e considerato un morfismo $t : T \to Y$, si ha (in termini di sottoinsiemi):

$$\begin{array}{ccc} S & & T \\ {\scriptstyle s}\downarrow & & \downarrow{\scriptstyle t} \\ X & \xrightarrow[f]{} & Y \end{array} \qquad S \subseteq f^*(T) \text{ se e solo se } \exists_f(S) \subseteq T.$$

Ma occorre esprimersi con piú esattezza. Si osservi a tal fine che i morfismi di codominio X possono considerarsi come gli oggetti di una nuova categoria $\mathbf{S}/X$ per la quale si possono definire i morfismi $\alpha : (S_1 \to X) \to (S_2 \to X)$ come triangoli commutativi

$$\begin{array}{ccc} S_1 & \xrightarrow{\alpha} & S_2 \\ & \searrow \quad \swarrow & \\ & X & \end{array}$$

Considerando dunque le due categorie $\mathbf{S}/X$ e $\mathbf{S}/Y$ è ora possibile ridescrivere f^* come un funtore, con $f^* : \mathbf{S}/Y \to \mathbf{S}/X$. Pensando ora esattamente $\exists_f(S)$ come $\exists_f(s)$, cioè considerando l'immagine di sf come un'azione sui morfismi e non sui sottoinsiemi, si ha un funtore $\exists_f$. Ora, come è immediato verificare,

$$\frac{s \to f^*(t)}{\exists_f(s) \to t},$$

cioè $\exists_f \dashv f^*$.

Questa definizione del quantificatore esistenziale, che ha come parallela la descrizione del quantificatore universale come $f^* \dashv \forall_f$, è evidentemente suscettibile di applicarsi a ogni situazione ove le costruzioni categoriali necessarie siano possibili, e dunque anche in contesti diversissimi dagli insiemi astratti.

Il caso dei fasci è molto istruttivo. Poiché piú volte in quest'*Enciclopedia* s'è definito rigorosamente il concetto di fascio (si veda ad esempio l'articolo «Differenziale», IV, p. 826), si presenterà qui una descrizione intuitiva che ha il merito tuttavia di suggerire le generalizzazioni presentate nel paragrafo seguente.

Si supponga di disporre di una serie di «stadi» U, U', U'', ..., per i quali si dia un concetto di «approfondimento». In termini tecnici, essi costituiscono un preordine. Si ha cioè una relazione tra le coppie U, U' che si indica con $U \geq U'$ riflessiva e transitiva: cioè 1) $U \geq U$; 2) $U \geq U'$ e $U' \geq U''$ implica

$U \geq U''$. Naturalmente, nel caso $U \geq U'$, si penserà U' come «uno stadio piú profondo».

Occorre ora assumere un'ulteriore proprietà. In termini intuitivi: dati due stadi U e V, esiste uno stadio che approfondisce entrambi $U \wedge V$. In termini rigorosi, si dirà che l'insieme preordinato ammette un estremo inferiore per ogni coppia di elementi.

Ecco ora il concetto di prefascio: un prefascio X è il dato di un insieme $X(U)$ per ogni stadio U e di un'applicazione tra insiemi $\rho_{UU'}$ per ogni coppia di insiemi $U \geq U'$, con le proprietà ulteriori: $\rho_{UU} = 1_{X(U)}$; $\rho_{UU'} \cdot \rho_{U'U''} = \rho_{UU''}$. Per continuare nella metafora, si può pensare a un elemento $x \in X(U)$ come a un «individuo allo stadio U». $\rho_{UU'}(x)$ sarà allora lo *stesso* individuo a uno stadio *ulteriore*.

È facile rendersi conto che tutti i dati di un prefascio sono racchiusi nell'osservazione che il preordine degli stadi U, U', ... è una particolare categoria $\mathbf{P}$, i cui oggetti sono gli stadi e i cui morfismi sono dati dalla corrispondenza $U \geq U' \mapsto U \leftarrow U'$. Un prefascio X è semplicemente un funtore controvariante $\mathbf{P}^{\mathrm{op}} \to \mathbf{S}$. Dati due prefasci X, Y, un morfismo tra prefasci, con tutta naturalità, è definito come una trasformazione naturale $f : X \dashrightarrow Y$. Questo significa che, per ogni $U \geq U'$, si ha il quadrato commutativo

$$\begin{array}{ccc} X(U) & \xrightarrow{f_U} & Y(U) \\ {\scriptstyle \rho_{UU'}}\downarrow & & \downarrow{\scriptstyle \sigma_{UU'}} \\ X(U') & \xrightarrow[f_{U'}]{} & Y(U') \end{array}$$

Naturalmente $\sigma_{UU'}$ sono le applicazioni corrispondenti di Y.

È facile interpretare la commutatività del quadrato: un elemento di $X(U)$ viene mandato da f_U in un elemento $f_U(x) \in Y(U)$. Ma $\rho_{UU'}(x)$ è «lo stesso elemento» allo stadio U', cosí come $\sigma_{UU'}(f_U(x))$ è «lo stesso elemento» allo stadio U'. Allora, ovviamente, chiedere la commutatività del quadrato precedente significa tener conto della particolare nozione di elemento/individuo che si è data.

Si immagini ora, in termini di prefasci, la situazione corrispondente a quella esaminata all'inizio di questo lungo esempio:

$$\begin{array}{ccc} S & & \\ {\scriptstyle s}\downarrow & & \\ X & \xrightarrow{f} & Y \end{array}$$

ove S, X, Y sono ora prefasci e s, f trasformazioni naturali. In questo caso $\exists_f$ e $\forall_f$ non sono suscettibili di alcuna interpretazione se non nei termini descritti, cioè come aggiunti sinistro e destro di f^*. E tali aggiunzioni si interpretano in modo molto suggestivo. Si veda ad esempio $\forall_f$, il piú interessante in questo caso. Occorre dunque che sia (con le stesse notazioni usate in precedenza)

$$\frac{t \to \forall_f(s)}{f^*(t) \to s}.$$

Come è facile verificare, questo conduce a descrivere $\forall_f(s)$ in questo modo: $\forall_f(s)=\{y \mid f^*(y)\subseteq s\}$. Se in particolare si pensa a $S\subseteq X$ e a $T\subseteq Y$ (senza badare al modo della inclusione) si ha $\forall_f(S)=\{y \mid f^*(y)\subseteq S\}$.

Ma le inclusioni che compaiono nella formula precedente sono, si ricordi, trasformazioni naturali, il che significa certo che per ogni livello U si debba avere $\forall_f(S)(U)=\{y\in Y(U) \mid f^*_U(y)\subseteq S(U)\}$, ma anche che deve esservi compatibilità tra i livelli. Non basta che y al livello U verifichi la definizione, occorre anche che per un livello V, con $U\geq V$, il trasformato di y, $\sigma_{UV}(y)$ verifichi la definizione al livello V. Cioè $\forall_f(S)(U)=\{y\in Y(U) \mid f^*_U(y)\subseteq S(U)$ e, per ogni $V\leq U$, $\sigma_{UV}(y)$ verifichi $f^*_V(\sigma_{UV}(y))\subseteq S(V)\}$. Si usa dire che l'interpretazione del quantificatore $\forall$ è universale nei prefasci. L'interpretazione del quantificatore $\exists$ è invece semplicemente quella livello per livello (attuale), e non è perciò molto interessante. Diviene invece di grande interesse nel caso dei fasci che si illustreranno ora.

Occorre un concetto di «ricoprimento» (topologia di Grothendieck) di uno stadio (se ne può trovare la definizione completa nell'articolo «Locale/globale» di questa stessa *Enciclopedia*, VIII, p. 486. Una trattazione dettagliata si ha poi in Johnstone [1977]). Per ogni stadio U sono assegnate certe famiglie $\{U_i \mid i\in I\}$ di sottostadi $U_i\leq U$, le quali si giudicano «coprenti» (la terminologia è presa dagli spazi topologici, dove queste famiglie sono i ricoprimenti aperti). In qualche modo, U è «rappresentato» dagli U_i. Naturalmente, per avere una teoria efficace occorre che queste famiglie siano descritte con opportuni assiomi (si veda ancora «Locale/globale»), ma, per le esigenze di quest'articolo, essi possono essere sottintesi.

Sia ora F un certo prefascio. Esso sarà specificato come un fascio descrivendo il suo comportamento rispetto alle famiglie coprenti. Sia $\{U_i\}$ un ricoprimento dello stadio U. Si hanno allora i morfismi ρ_{UU_i}, i quali associano a ogni individuo allo stadio U una collezione di individui x_i agli stadi U_i, con $x_i=\rho_{UU_i}(x)$. Si ottiene un fascio quando si specifica quali sono le famiglie x_i di elementi di $F(U_i)$ che provengono esattamente in questo modo. Se U_i e U_j sono due sottostadi di U del ricoprimento, l'ipotesi fatta sulla relazione d'ordine è che esista $U_i\wedge U_j$. Si hanno allora anche i morfismi $\rho_{U_iU_i\wedge U_j}$, $\rho_{U_jU_j\wedge U_i}$, i quali consentono di trasportare un elemento di U_i e un elemento di U_j al livello $U_i\wedge U_j$. Si ha allora il concetto di famiglia compatibile: una famiglia di elementi $x_i\in F(U_i)$ tale che $\rho_{U_iU_i\cap U_j}(x_i)=\rho_{U_jU_i\cap U_j}(x_j)$. Ed ecco ora la

DEFINIZIONE. *Il prefascio F è un fascio se per ogni famiglia compatibile x_i come quella precedente esiste esattamente un $x\in F(U)$ con $\rho_{UU_i}(x)=x_i$.*

La situazione

$$\begin{array}{ccc} S & & \\ {\scriptstyle s}\downarrow & & \\ X & \xrightarrow{f} & Y \end{array}$$

può ora essere esaminata in termini di fasci. Si vede l'interpretazione di $\exists_f(s)$. O meglio, semplificando ancora, con l'inclusione $S\subseteq X$, $\exists_f(S)$. Il risultato è il

seguente: $y\in\exists_f(S)(U)$ non solo quando esiste un x in $S(U)$ con $f_U(x)=y$, ma anche quando ciò avviene per un numero sufficiente di livelli ulteriori. Esattamente, esiste un ricoprimento $\{U_i\}$ tale che gli $y_i=\rho_{UU_i}(y)$ sono in $f_{U_i}(S(U_i))$.

L'elemento y può non stare nell'immagine a un certo livello, ma questo può succedere a y in stadi «piú profondi». Se questo avviene un numero sufficiente di volte (per un ricoprimento) allora y è in $\exists_f(S)(U)$. Si dice anche che y è «localmente» nell'immagine.

Queste profonde idee di Lawvere possono essere descritte in molti ulteriori e fecondi sviluppi sui quali non ci si soffermerà. Ci si limiterà ad osservare che l'interpretazione dei quantificatori nei fasci apre la via a una trattazione della «logica» in questa situazione che è ricca di sviluppi profondi. [Per un'esposizione esauriente e dettagliata si veda Meloni 1974-77].

5. *Il concetto di topos elementare.*

Il concetto di topos di Grothendieck è già intervenuto in numerosi articoli di questa *stessa* Enciclopedia (si veda ancora «Locale/globale», VIII, pp. 486-487) e già s'è accennato all'idea di Lawvere di considerare ogni topos come l'ambito naturale di una «logica interna» ad esso canonicamente associata. Tuttavia l'idea di Lawvere è anche intimamente connessa a una generalizzazione del concetto di topos di Grothendieck, senza la quale probabilmente essa sarebbe assai meno efficace.

Per citare Johnstone, l'idea scaturisce dall'osservare che «ogni topos di Grothendieck ha un oggetto dei valori di verità Ω, e che la nozione di topologia di Grothendieck è strettamente connessa agli endomorfismi di Ω» [1977, p. XIV]. Qui ci si limiterà a presentare il primo punto, rinviando allo stesso Johnstone [*ibid.*, pp. 76 sgg.] per il concetto di topologia in un topos elementare e per altri interessanti sviluppi della teoria. Si vedrà nel paragrafo successivo un'importante direzione di ricerca non ancora oggetto di sviluppi sistematici.

DEFINIZIONE. *Un topos elementare è dato da una categoria* **E** *1) con tutti i limiti finiti; 2) che ammetta per ogni oggetto X un funtore esponenziale $(\ )^X$ aggiunto destro del funtore $(\ )\times X$. (In tal caso,* **E** *si dice cartesianamente chiusa); 3) che ha un sotto-oggetto classificatore Ω; cioè un oggetto e un morfismo «vero» $v: 1\to\Omega$, tale che per ogni monomorfismo $f: Y\to X$ vi è una funzione classificante φ_f, con $\varphi_f: X\to\Omega$ tale che sia un prodotto fibrato il diagramma*

$$\begin{array}{ccc} Y & \longrightarrow & 1 \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle v} \\ X & \xrightarrow{\varphi_f} & \Omega \end{array}$$

L'assioma piú caratteristico è certamente il terzo; se ne veda ora l'interpretazione nel caso degli insiemi astratti. Ω è dato semplicemente da $\{0,1\}$ (falso e vero) e v è definita $v(1)=1$. Un monomorfismo f è un «modo» per descrivere Y come sottoinsieme di X. φ_f è la funzione caratteristica, che vale 1 sugli ele-

menti di X che sono immagini di elementi di Y e o altrimenti. Nel caso generale Ω non si ridurrà a due elementi, ma conterrà altri «valori di verità».

Come si vede, tutto è assai semplice, in questo caso. Ciò dipende da una contraddizione (sottolineata da Lawvere) che si manifesta nel caso degli insiemi: da un lato gli insiemi, come il caso piú «astratto», rappresentano una situazione poco interessante (si tenga presente che gli insiemi possono pensarsi come i fasci sulla categoria formata da un solo oggetto con i morfismi ridotti alla identità); d'altro lato, quando viene colta negli aspetti essenziali una costruzione che può essere effettuata negli insiemi – con attenzione ad utilizzare come strumenti dimostrativi solo quelli della logica intuizionista –, la costruzione può essere trasferita in ogni topos. Gli insiemi astratti sono dunque un formidabile strumento euristico.

Un teorema [se ne veda la dimostrazione *ibid.*, p. 24] e un esempio daranno ora un'idea dell'estensione del concetto di topos elementare:

TEOREMA. *Ogni topos di Grothendieck è un topos elementare.*

D'altro lato, come è assai facile verificare, la categoria degli insiemi finiti non è un topos di Grothendieck, pur essendo un topos elementare.

I topoi elementari sono dunque «piú generali» dei topoi di Grothendieck, il che, come ricorda l'epigrafe posto all'inizio di questo articolo, può essere di interesse assai scarso. Ma il fatto è che essi colgono, con la descrizione assiomatica che s'è vista, gli aspetti *realmente centrali* della teoria, consentendone uno sviluppo facile e diretto. La maggior generalità fornisce al tempo stesso un prezioso materiale per interpretare teorie logiche prive di modelli nel caso classico. Nel paragrafo seguente ci si soffermerà su un esempio di questo genere; non si accennerà invece agli sviluppi della teoria, per cui si può vedere Meloni [1974-77].

6. *Il calcolo differenziale dal punto di vista categoriale.*

In occasione della pubblicazione degli atti di un convegno su *Topos-Theoretic Methods in Geometry and Analysis*, tenuto ad Aarhus nel maggio 1978, il curatore Kock ha posto come primo articolo *Categorical Dynamics* di Lawvere. Si tratta di una serie di conferenze che risalgono al 1967, le quali, pur rimaste inedite per anni, hanno avuto larga influenza su molti ricercatori. In una nota esplicativa posta all'inizio, Lawvere osserva come: «In queste conferenze, io davo alcuni calcoli preliminari per favorire un programma per (3) assiomatizzare la fondazione della meccanica dei continui nello spirito di Walter Noll sulla base di (2) una assiomatizzazione diretta dell'essenza della topologia differenziale usando risultati e metodi dei lavori francesi in geometria algebrica (alcuni dei quali ho appreso da Gabriel); ma mi accorsi poi che questo richiede (1) lo studio assiomatico delle categorie di insiemi lisci, simili ai topoi di Grothendieck, poiché la forma piú naturale di (2) è incompatibile con la usuale teoria degli insiemi» [1967, p. 1].

Ci si soffermerà brevemente sul punto (2) indicando uno sviluppo dovuto essenzialmente a Kock, utilizzando il primo [Kock 1975-76] di una serie di articoli su questo problema.

Cosa significa intanto l'incompatibilità con l'usuale teoria degli insiemi? Significa essenzialmente che in essa non è possibile disporre di un'adeguata formalizzazione dei procedimenti differenziali con l'uso diretto degli infinitesimi. Ciò è invece possibile per gli «insiemi variabili», i topoi, e si ottiene cosí il duplice vantaggio di utilizzare strutture spesso piú «aderenti» a ciò che si esamina – i topoi appunto –, ove i procedimenti differenziali si dànno in modo assai semplice.

L'idea-guida di quanto si esporrà ora è che per una funzione f, definita in un valore a, se d è un infinitesimo, sarà $f(a+d)=f(a)+b\cdot d$. Cioè una funzione lineare. Ovviamente allora $b=f'(a)$.

Come si vede non è esattamente una novità, da un certo punto di vista: lo è invece nel senso che esistono contesti non banali ove questo approccio è possibile e agevole.

Sia allora **E** un topos elementare (in realtà per molti sviluppi si può chiedere meno, ma ci si pone direttamente in questo contesto per semplicità) e sia A un oggetto anello in esso. Ciò significa che sono dati certi morfismi $A\times A\xrightarrow{+}A$, $A\times A\xrightarrow{\cdot}A$, $A\xrightarrow{-(\ )}A$, $1\xrightarrow{\ulcorner 0\urcorner}A$, $1\xrightarrow{\ulcorner 1\urcorner}A$, che corrispondono alle abituali operazioni di un anello, la somma, il prodotto, ecc. Naturalmente queste operazioni debbono soddisfare a determinati assiomi, il che significa che si richiede la commutatività di certi diagrammi. Si vede, ad esempio $a+0=a$, per ogni elemento a. Ciò significa che i due morfismi

$$\begin{tikzcd} A \simeq A\times 1 \arrow[r, "\mathrm{id}\times 0"] \arrow[d, equal] & A\times A \arrow[r, "+"] & A \\ A \arrow[urr, "1_A"'] & & \end{tikzcd}$$

debbono essere uguali. In modo analogo si procede per gli altri assiomi.

Occorre ora l'oggetto D degli infinitesimi. Esso è fornito dall'ugualizzatore dei due morfismi

$$\begin{tikzcd} A \arrow[r, "\langle \mathrm{id}, \mathrm{id}\rangle"] \arrow[dr] & A\times A \arrow[r, "\cdot"] & A \\ & 1 \arrow[ur, "\ulcorner 0\urcorner"'] & \end{tikzcd}$$

In termini di «elementi» si tratta dei d tali che $d^2=0$. Accanto a D si hanno i due funtori $(\ )\times D$ e $(\ )^D$, con $(\ )\times D\dashv(\ )^D$ (cfr. p. 509), e dunque i morfismi

$$A\times D\xrightarrow{+}A$$
$$A\xrightarrow{\hat{+}}A^D$$

ove $\hat{+}$ si ottiene per aggiunzione. Ancora, in termini di elementi, $\hat{+}$ è il morfismo definito da $a\mapsto a+(\ )$ ove la parentesi va «riempita» con gli infinitesimi. $a+(\ )$ è la funzione che per ogni d vale $a+d$.

Sia ora α il morfismo $A \times A \xrightarrow{\alpha} A^D$ che, descritto direttamente sugli elementi, è dato da $(a_1, a_2) \mapsto a_1 + a_2(\ )$ [cfr. *ibid.*, p. 3 per la descrizione in termini di diagrammi].

Ora è facile osservare che A^D, come nel caso degli insiemi ove si definiscono le operazioni «punto per punto», è a sua volta un anello; α si mostra facilmente essere un omomorfismo di anelli. Ed ecco ora il fatto fondamentale [cfr. *ibid.*, p. 8]:

ASSIOMA. *Il morfismo α è invertibile.*

Sia ora γ il morfismo definito da $A^D \xrightarrow{\alpha^{-1}} A \times A \xrightarrow{p_2} A$, ove p_2 è la seconda proiezione. Ed ecco ora l'operazione di derivazione:

DEFINIZIONE. *Dato il morfismo $f: A \to A$, la sua derivata è data da*

$$A \xrightarrow{\hat{+}} A^D \xrightarrow{f^D} A^D \xrightarrow{\gamma} A.$$

Si osserverà che $A \xrightarrow{\hat{+}} A^D \xrightarrow{f^D} A^D$ ha per aggiunta $A \times D \xrightarrow{+} A \xrightarrow{f} A$, che è il morfismo definito sugli elementi con $(a, d) \mapsto f(a+d)$. Allora, per il significato dell'aggiunzione, il morfismo $A \xrightarrow{\hat{+}} A^D \xrightarrow{f^D} A^D$ è dato, ancora sugli elementi da $f(a+ + (\ ))$. Ora, con γ, prima si associa a questa funzione l'unica funzione lineare corrispondente (per l'assioma) – sia $a_1 + a_2(\ )$ – e poi con la seconda proiezione si prende a_2.

Questa nella sostanza la spiegazione della definizione. Da questa definizione conseguono con grande facilità tutti i consueti sviluppi del calcolo differenziale sui quali, ovviamente, non è il caso di insistere.

Ed ecco ancora un possibile modello [cfr. *ibid.*, pp. 13-14]. Sia **R** la categoria degli anelli commutativi finitamente presentati, e sia ζ la categoria dei funtori covarianti verso la categoria degli insiemi. Allora l'inclusione di Yoneda $\mathbf{R}^{op} \xrightarrow{y} \zeta$ valutata su $Z[X]$ dà un oggetto anello con le proprietà richieste. Si tratta naturalmente del «riflesso» della abituale differenziazione formale dei polinomi [*ibid.*, p. 17].

Il lettore avrà osservato come spesso accanto ai diagrammi v'era una facile «lettura» in termini di elementi. Non è possibile perseguire *direttamente* questo punto di vista? In effetti molti degli ultimi approcci stimolati ancora da Lawvere vanno in questa direzione. Naturalmente occorre però rendersi conto che, poiché i modelli non sono negli insiemi ma negli «insiemi variabili», occorre trattare gli elementi con la logica intrinseca ad essi, cioè la logica intuizionista [cfr. Casella e Meloni 1980; Meloni 1974-77].

7. *Conclusione.*

All'autore di quest'articolo è capitato, chiedendo a Takeuti un giudizio su una recente ridimostrazione, da parte di Lawvere e Tierney, del risultato di Cohen sull'ipotesi del continuo, d'avere nella sostanza questa risposta: con la

teoria delle categorie si può riformulare, chiarire, penetrare piú a fondo se si vuole; ma non si può in realtà «dimostrare». I suoi metodi sono troppo astratti e generali per intervenire con reale efficacia a promuovere risultati. Solo *dopo* che un risultato è stato acquisito si può intervenire in maniera categoriale.

È certo importante, si può osservare, che il lemma di Yoneda unifichi cose all'apparenza assai distanti come le sezioni di Dedekind e il teorema di rappresentazione di Cayley, ma questo è un fatto «concettuale», non un effettivo risultato suscettibile di incrementare lo sviluppo della ricerca matematica.

A obiezioni di questo genere si potrebbe rispondere citando titoli di merito, teoremi effettivamente dimostrati, ecc. Ma questa risposta sarebbe estrinseca. Si può invece farsi carico delle obiezioni e vedere che cosa esse realmente significano. Che cosa, in effetti? Riflettono una concezione della matematica che tiene separati da un lato l'operare «concreto», l'effettivo ottenimento dei risultati, dall'altro il riflettere sui metodi, sulle procedure sui legami tra le parti della matematica. Separazione forzata e insostenibile.

Hilbert, a detta di Gordan, avrebbe fatto teologia e non matematica sciogliendo il famoso problema degli invarianti, ma chi oggi non accetta, oltre ai suoi risultati, anche le tecniche usate?

È certo vero che l'aritmetica di Frege non aiuta affatto ad eseguire meglio le operazioni aritmetiche, ma è difficile sostenere che essa non abbia avuto esiti importanti anche sulla «pratica» matematica: si pensi all'importanza dell'analisi delle dimostrazioni dal punto di vista del calcolo con l'elaboratore. Cosí ciò che realmente viene a manifestarsi non è che la volontà di tenere separato ciò che per sua natura è intimamente unito.

Infine un'osservazione: soprattutto a partire dagli anni '70, ossia dai topoi elementari, s'è avuto uno sviluppo imponente della ricerca nel campo delle categorie: si hanno ormai molte centinaia di articoli. Difficile cosí dare una visione sistematica e unitaria, enucleare le idee centrali. Ma, in ogni caso, è questo che occorre fare, e che certo quest'articolo avrà fatto solo in parte, rifiutando la comoda e sgradevole immagine d'un vasto inesplorato e inesplorabile territorio ove l'«esperto» non può che manifestare la propria solidarietà al «profano», consacrando cosí il proprio ruolo di funzionario. [M. G.].

Bergson, H.-L.
1907 *L'Evolution créatrice*, Alcan, Paris; ed. Presses Universitaires de France, Paris 1970.

Casella, A., e Meloni, G. C.
1980 *Introduzione alle quantità variabili con infinitesimi*, Istituto di Matematica dell'Università di Milano, Milano.

Eilenberg, S., e MacLane, S.
1942 *Natural isomorphisms in group theory*, in «Proceedings of the National Academy of Sciences of the United States of America», XXVIII, pp. 537-43.
1945 *General theory of natural equivalences*, in «Transactions of the American Mathematical Society», LVIII, pp. 231-94.

Frajese, A., e Maccioni, L.
1970 Commento agli *Elementi di Euclide*, Utet, Torino.

Johnstone, P. T.
1977 *Topos Theory*, Academic Press, New York.

Kan, D. M.
1958 *Adjoint functors*, in « Transactions of the American Mathematical Society », LXXXVII, pp. 294-329.

Kant, I.
1787 *Kritik der reinen Vernunft*, Hartknoch, Riga 1787[2] (trad. it. Utet, Torino 1967).

Kock, A.
1975-76 *A Simple Axiomatic for Differentiation*, Matematisk Institut, Aarhus Universitet, Aarhus.

Lawvere, W. F.
[1967] *Categorical Dynamics*, in A. Kock (a cura di), *Topos-Theoretic Methods in Geometry and Analysis*, Matematisk Institut, Aarhus Universitet, Aarhus 1978.
1969 *Adjointness in foundations*, in « Dialectica », XXIII, 3-4, pp. 281-95.
1972-73 *Teoria delle categorie sopra un topos di base*, Istituto di Matematica dell'Università di Perugia, Perugia.

Lolli, G.
1977 *Categorie, universi e principî di riflessione*, Boringhieri, Torino.

MacLane, S.
1971 *Categories for the Working Mathematician*, Springer, New York - Heidelberg - Berlin (trad. it. Boringhieri, Torino 1977).

Meloni, G. C.
1974-77 *La logica degli insiemi variabili con continuità: algebra intuizionista e teoria dei fasci*, Istituto di Matematica dell'Università di Milano, Milano.

Vico, G.
[1725-28] *Vita di Giambattista Vico scritta da se medesimo*, s. e., Venezia 1729; ed. Einaudi, Torino 1977.

Al concetto di categoria si è giunti dopo che il **concetto** di trasformazione naturale è stato astratto (cfr. **astratto/concreto**) dalla pratica della topologia algebrica (cfr. **geometria e topologia**). Ma esso è anche al termine di un processo (cfr. **dialettica**, **storia**, **tempo/temporalità**) non privo di interazioni con il pensiero filosofico (cfr. **categorie/categorizzazione**, **filosofia/filosofie**, **sistematica e classificazione**) e con le riflessioni della fine del secolo scorso sulla natura della geometria (cfr. **invariante**). La teoria (cfr. **teoria/modello**) delle categorie, limitata dapprima a un comodo **linguaggio** per descrivere certe realtà (cfr. **reale**, **spiegazione**) **matematiche**, ha acquisito due importanti approfondimenti con la scoperta del concetto di aggiunzione e soprattutto con il concetto di topos (cfr. **applicazioni**, **strutture matematiche**), proseguito poi con quello di topos elementare (cfr. **elementi**, **infinitesimale**, **locale/globale**). Di qui importanti interazioni con la **logica**, la geometria algebrica e, in generale, il dispiegarsi di un vasto **programma** fondazionale.